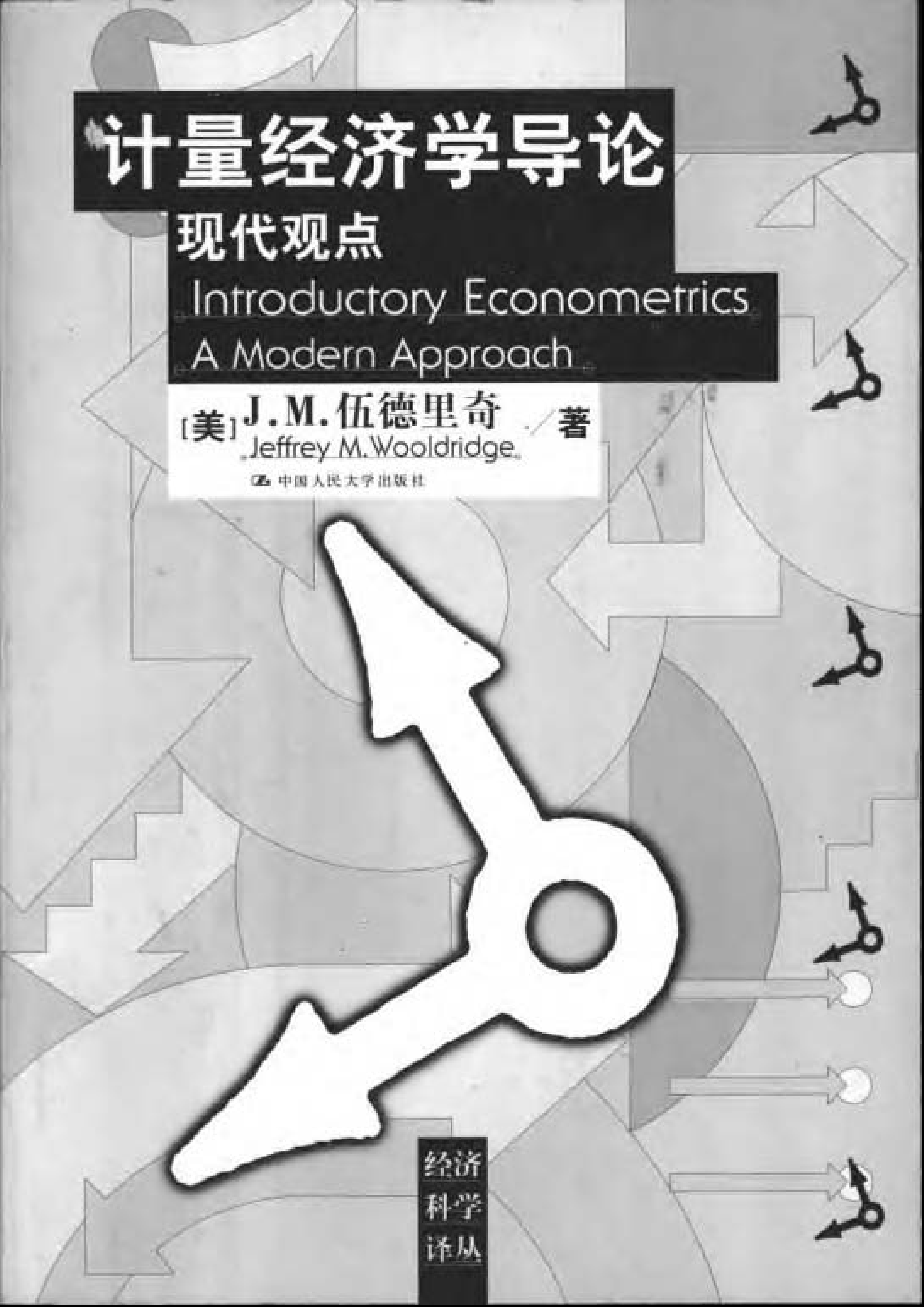


计量经济学导论

现代观点

Introductory Econometrics

A Modern Approach

[美] J.M.伍德里奇 著

Jeffrey M. Wooldridge

中国人民大学出版社

经济科学译丛

著作权合同登记号
图字：01-2003-1206号

计量经济学导论
现代观点
Introductory Econometrics
A Modern Approach

本书的面世恰逢时机，对应用研究者时常面临的几个重大问题，作者的系统处理方法已获得极大赞誉：

◎“在初级计量经济学教科书中，只有本书对时间序列数据的计量分析进行了认真而又圆满的讨论……它无须过分严密的推导而对复杂的计量经济思想进行了清晰而又直观的表述。”

——克里斯·康韦尔　佐治亚州立大学

◎“……伍德里奇完成了一件出色的工作，许多基本计量经济学概念和问题时常都迷失在错综复杂的数学公式之中，但本书总能给出非常直观的解释，因此，我建议所有的学生和应用计量经济方面的研究者们都把此书作为常备的参考书。”

——约翰·卡姆　纽约州立大学宾厄姆顿分校

◎“我曾以为市场上不可能还有另一本计量经济学教科书的立足之地，但我吃惊而又兴奋地看到这种看法被证明是错误的，伍德里奇的书采用了全新的方法，其对现代实践、综列数据和工具变量的强调都颇有见地。”

——爱德华·科尔森　宾夕法尼亚州立大学

◎“让第一学期的计量经济学课程就与实证研究齐头并进的时刻到了，我完全同意，不用微积分和太多的数学，也能培养出一流的实证计量经济学家。本书与我过去几年曾讲授的一门深受学生喜欢的课程十分相似。”

——皮特·科克伦　哈佛大学

榮晶工作室

ISBN 7-300-04518-9
9 787300 045184 >

ISBN 7-300-04518-9/F · 1387
定价：85.00元

梁晶工作室

计量经济学导论

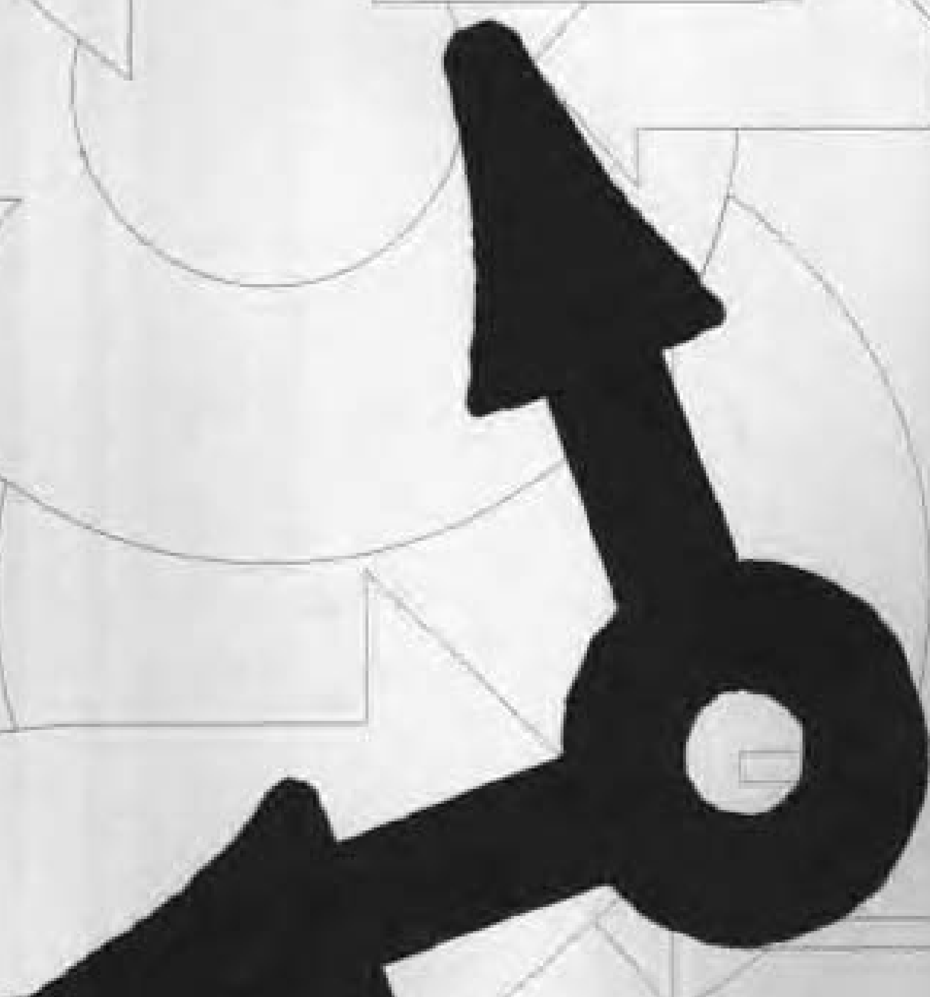
现代观点

Introductory Econometrics

A Modern Approach

[美] J.M.伍德里奇 著
Jeffrey M. Wooldridge

林少宫 / 校
费剑平 林相森 / 译



经济科学译丛

中国人民大学出版社

图书在版编目（CIP）数据

计量经济学导论：现代观点／（美）伍德里奇著；费剑平，林相森译
北京：中国人民大学出版社，2003
（经济科学译丛）

ISBN 7-300-04518-9/F·1387
Ⅰ．计…
Ⅱ．①伍…②费…③林…
Ⅲ．计量经济学
Ⅳ．F224.0

中国版本图书馆 CIP 数据核字（2003）第 002505 号

经济科学译丛
计量经济学导论：现代观点
［美］J.M. 伍德里奇　著
林少宫　校
费剑平　林相森　译

出版发行：中国人民大学出版社
（北京中关村大街 31 号　邮编 100080）
邮购部：62515351　门市部：62514148
总编室：62511242　出版部：62511239
本社网址：www.crup.com.cn
人大教研网：www.ttrnet.com
经　　销：新华书店
印　　刷：涿州市星河印刷厂

开本：787×1092 毫米 1/16　印张：53.5 插页 5
2003 年 3 月第 1 版　2003 年 3 月第 1 次印刷
字数：1 121 000

定价：85.00 元

（图书出现印装问题，本社负责调换）

《经济科学译丛》总序

中国是一个文明古国，有着几千年的辉煌历史。近百年来，中国又由盛而衰，一度成为世界上最贫穷、落后的国家之一。1949年中国共产党领导的革命，把中国从饥饿、贫困、被欺侮、被奴役的境地中解放出来。1978年以来的改革开放，使中国真正走上了通向繁荣富强的道路。

中国改革开放的目标是建立一个有效的社会主义市场经济体制，加速发展经济，提高人民生活水平。但是，要完成这一历史使命决非易事，我们不仅需要从自己的实践中总结教训，也要从别人的实践中获取经验，还要用理论来指导我们的改革。市场经济虽然对我们这个共和国来说是全新的，但市场经济的运行在发达国家已有几百年的历史，市场经济的理论亦在不断发展完善，并形成了一个现代经济学理论体系。虽然许多经济学名著出于西方学者之手，研究的是西方国家的经济问题，但他们归纳出来的许多经济学理论反映的是人类社会的普遍行为，这些理论是全人类的共同财富。要想迅速稳定地改革和发展我国的经济，我们必须学习和借鉴世界各国包括西方国家在内的先进经济学的理论与知识。

本着这一目的，我们组织翻译了这套经济学教科书系列。这套译丛的特点是：第一，全面系统。除了经济学、宏观经济学、微观经济学等基本原理之外，这套译丛还包括了产业组织理论、国际经济学、发展经济学、货币金融学、公共财政、劳动经济学、计量经济学等重要领域。第二，简明通俗。与经济学的经典名著不同，这套丛书都是国外大学通用的经济学教科书，大部分都已发行了几版或十几版。作者尽可能地用简明通俗的语言来阐述深奥的经济学原理，并附有案例与习题，对于初学者来说，更容易理解与掌握。

经济学是一门社会科学，许多基本原理的应用受各种不同的社会、政治或经济体制的影响，许多经济学理论是建立在一定的假设条件上的，假设条件不同，结论也就不一定成立。因此，正确理解掌握经济分析的方法而不是生搬硬套某些不同条件下产生的结论，才是我们学习当代经济学的正确方法。

本套译丛于1995年春由中国人民大学出版社发起筹备并成立了由许多经济学专家学者组织的编辑委员会。中国留美经济学会的许多学者参与了原著的推荐工作。中国人民大学出版社向所有原著的出版社购买了翻译版权。北京大学、中国人民大学、复旦大学以及中国社会科学院的许多专家教授参与了翻译工作。在中国经济体制转轨的历史时期，我们把这套译丛献给读者，希望为中国经济的深入改革与发展做出贡献。

《经济科学译丛》编辑委员会
1996年12月

前 言

在初等计量经济学的教学方法和经验（实证）研究者所想像、所应用，以至所解释的计量经济学方法之间，存在着越来越大的差距。其部分原因是，现代的经验研究者在某种程度上有赖于计量经济学的新进展，而这些进展仅缓慢地出现在教科书上。不过，许多经验研究所依赖的，也仅是初等教程中所讲授的比较小的一部分计量经济学工具而已。

大多数教科书在陈述和解释其假定时，采取了图方便而忽视现实的选择，使得学生在阅读和从事认真的经验研究方面缺少应有的准备。这种现实性的缺乏，反映在充斥着大多数现有计量经济学入门教科书中过于简单化或处理不当的例解中。

在编写本书时，我希望能为计量经济学的入门教学重新取向，哪怕是很有限度的。我相信，计量经济学应被人们使用，而不是被动地加以吸收。学生们最好的学习方法是，在他们剖视一位应用经济学家的时候，以及在他们学习代表着现代实践的许多经验性应用的时候，学会怎样使用计量经济学。

本书和已习用多时的教材相比，有着明显的差别。然而我相信，大多数授课的教师，特别是带有应用倾向的教师，将发觉我对计量经济学的处理方法比现行的方法更为亲切。我听惯了这样一种说法：想教会学生沿着当前计量经济学实践的方式，利用计量经济学来回答重要的有趣问题是徒劳的。按照这种说法，充其量，你只能提供给学生一个计量经济学概貌，然后，通过

一些呆板的例解，机械地阐述其基本方法。我不同意这种观点。大学本科生不仅能学会反映着现代实践的计量经济学，而且这样的教学观点还会增进他们对学习计量经济学的兴趣和享受。

我的处理方法中的一些特点可追溯到我最早期的计量经济学作品。与其放弃早期的方法或试图包揽一大堆计量经济学方法，我宁可选择对阅读期刊文献和从事基本经验研究有用的专题。对于每一专题，我有意略去许多经不起时间考验的检验方法和步骤，尽管传统教科书都要把这些包括进来。同时，我把已表明有明显用途的较新的专题放进书中，诸如导出对未知形式的异方差（或序列相关）保持着稳健性的检验统计量，利用多年数据进行政策分析，或通过工具变量法解决遗漏变量问题。

我想我的处理方法既是系统的，又是统一的。所谓系统，指的是每一专题的叙述都合乎逻辑地建立在先前的材料之上，并且每一假定都是为了得到某一结论所必需才引入的。例如，应用研究者和理论家都知道，为了证明普通最小二乘法的无偏性，并不需要全部高斯-马尔科夫假定。然而，几乎所有计量经济学教本在证明最小二乘法的无偏性之前，就把全部假定引了进来，这样做只能给学生造成混乱。

在本书中，仅当必要时才把假定引进来，这样做便于对每一假定的含义进行仔细的、直觉的讨论：为什么该假定是必要的。同时，这种系统方法由于去掉了多余的假定而显得简洁。我的系统方法还用于研究大样本性质和更高深的方法，诸如混合横截面、纵列数据和工具变量法的利用。我的方法在下述意义上是统一的：所有的估计量和检验统计量均可利用少数的直觉而合理的估计和检验原理而获得（当然，也都可以严格推理得到）。这和动辄给出表面看来无关的一组“计量经济学步骤”处方的处理方法形成了对照。

现在通用的计量经济学方法的大部分均可从少数的基本估计原理推出，这并非偶然现象。这样一来确实减轻了学生的负担，同时不影响他们对获得的结果有一个牢靠的理解，而不至于把这些结果看成什么神秘的计量经济学方法。

我对多元回归分析的处理方法，是把计量经济模型中误差或干扰的性质的讨论同不可缺少的条件期望这一工具结合起来。不同于其他教材，在陈述和解释假定时，我完全放弃了非随机的或在重复样本中加以固定的回归元假定。在诸如经济学、政治学、社会学、城市研究、教育学、会计学、金融学和市场营销等使用非实验数据的学科领域里，熟练的应用研究者不会按照固定了的回归元那样思考问题，因为那是不真实的和把人引入歧途的。反而重要的是，观测不到的误差和观测到的解释变量究竟有什么关系，这是贯穿全书的讨论焦点。

这种方法有利于驱除出现在各个等级的计量经济学教材中的若干误解。例如，当出现滞后应变量时，在什么情况下普通最小二乘法是一致性的？在出现异方差或序列相关时，用通常的 R-平方作为拟合优度的一种度量，为什么仍然生效？或者，为什么不应把对函数形式的检验看做对遗漏的、观测不到的变量的一般检验？

通观全书，我一直强调所谓其他条件不变的含义，这说明我在只用一章的篇幅讨论简单的回归模型之后便进入多元回归分析，目的是要学生尽早地认真考虑实际应用。比起现有的教材，我对各种数据结构的政策分析要重视得多。对一些实用专题，诸如利用代理变量以达到其他条件不变的效果，以及对含有交互作用项的模型求其偏效应的标准误，也都作了简要的处理。

本书含有大量例题，许多是取自或受启发于应用经济学或其他领域的最新及有影响的作品。其中，大多数的数值计算都可以利用书中所附的数据集加以验证。我收集这些数据集已经多年了，发现这些数据迎合了学生的兴趣。

听众对象

本教材针对那些修过一门初等概率与统计的经济学和企业专业的大学本科生而写。多元回归分析方面的基本材料，由于本书中所强调的是解释和例解，故适合于大多数本科经济系作为一个入门课程内容而讲授。可根据学生的能力和兴趣掺进公式推导。某些专题，虽然略为高深，但可作为一个入门课程的后续部分，或者放在一个高年级讨论班上作为启发创始性经验研究来安排。

本教材也很适合作为硕士研究生水平的一个导引课程，但讲课的进程要快些，公式推导要多些（可能的话，包括使用一些矩阵代数）。某些较高深的专题，包括纵列数据方法、工具变量法、限值应变量以及时间序列计量经济学的新近进展，都可在这一水平上讲授。

本教材对所有应用领域的博士研究生都会是有用的，特别是在陈述假定时所采取的途径和方法上。本书和研究生用的教科书相比，把第三部分的较高深方法浅化了许多许多。对工具变量和纵列数据方法的处理，特别着眼于来自劳工经济学、城市经济学、公共财政、国际经济学以及政策分析方面的新近应用。

教材组织特点

本教材和其他教材的最明显的组织差异也许是：本教材的第一部分专门讲使用横截面数据的回归分析，然后在第二部分讲使用时间序列数据的回归。微观经济领域的应用研究者大概会发现这样做是有意义的。我所以做这样的选择有几个理由：最为重要的是，教学得以顺利进行，而且学生感到很自然。在横截面数据来自随机抽样的假定下，讨论的重点可完全放在总体模型和对总体的假定上。（在课程的以后部分可轻易地以直观的方式开展关于非随机抽样的后果的讨论。）把焦点摆在总体模型上，使得随机解释变量的引入异常简单，并强调了适当陈述的零均值和同方差性假定是以解释变量为

条件的。

由于这些假定比较明显而现实，我的方法能使学生早日从事于认真的横截面应用，而无须顾虑时间序列回归模型中可能出现的趋势、动态、高度持久性和谬误回归等麻烦问题。

教学计划

作为一学期课程，虽然可以做不同的教学安排，但我建议，应包括第1～8章、第9章的一部分（主要通过例题讲授）以及第10～12章。根据你的判断，可以跳过一些不常见的材料，诸如对序列相关而言的普通最小二乘法稳健性推断和动态异方差模型。

如时间允许，可讲些第三部分的材料。这些材料可以为第二学期课程打好基础。从教学上看，多数专题都不会比第一、二两部分难多少。

其中一些涉及更深奥的概念而不是深奥的数学。第三部分的一个重要特点，是对特殊数据结构，主要是不同时间的独立横截面的混合和纵列数据作了凸显的考虑。第13章是特别易于理解的，它讨论了用自然实验的数据作政策分析和项目评估等重要议题。

第15、16章，关于工具变量估计法和联立方程模型，对经验研究中所遇到的重要数据结构类型，均一一加以考虑。第15章把工具变量作为解决遗漏变量问题的一种方法而引进。被遗漏的变量问题比第16章的联立性问题要更为常见而且在概念上也要简单得多。

第17章是惟一的考虑模型中参数为固有非线性的一章，当学生做应用微观计量经济学方面的研究论文时，可作为原始资料。本章在初级水平上讨论了二值响应的概率单位和对数单位模型以及计数响应的泊松回归。我明确表示，当数据的收集过程对应变量实行截取时，要利用截取回归，而托比模型则适用于求角点解结果。

第18章涉及时间序列计量经济学近来的一些重要专题，如单位根和协积的检验。本章还包含有一个相当详细的关于预测的导论。

第19章是为了那些要求写期末论文的课程而附加到授课提纲里来的。同其他书本中的类似章节相比，其内容要显得广泛得多；它概括了适合于各种问题、各种数据结构的一些方法，指出潜在的陷阱，较详细地解释怎样写一篇经验经济学的期末论文，并提出可能研究的项目。

本书附有广泛的附录，包括基础代数、概率、统计和用于多元回归的矩阵代数方法。附录A复习了学生们在选读计量经济学就应懂得的数学工具，特别是对总和运算子、比例、百分数、对数和其他重要函数作了细致的讨论。附录B和C，虽然无意用以替代一门概率与统计课程，但其自身却是相当完备的。一位有专业旨趣的学生不妨把它们连同附录A一起读完，这样就可以为阅读正文做良好的准备。

考虑到一些人的偏好，我还对用矩阵讲授多元线性回归写了附录D和

E。但是，为了赏识各种计量经济学方法的应用性，或理解各种假定和结果，这些材料并非必需的。

设计特色

本书有几个特色使它成为学生和讲师们的良师益友。每章都含有一些边学边问的问题，并在附录中给出了答案；这些问题的目的在于督促学生及时反思。每章也都含有许多编号的例题，其中的一些取自最近发表的论文中的案例研究，但都根据我的判断作了分析上的简化，无损于其要点。

章末的习题和计算机作业题侧重于经验（实证）习作，而不是复杂的推导。要求学生根据他们所学，仔细地加以理解。计算机作业往往是课文中的例题加以扩充，若干习题利用了已出版作品中的数据集或由经济学及其他领域中发表过的研究成果所启发的类似数据集。

本书利用了ASCII版本的多于60个数据集。因为许多数据集来自实际研究工作，其中一些异常庞大，所以除了为说明各种数据结构而部分地列出数据集外，书中都不把数据集报导出来。本书的设计适合于计算机作业扮演着重要角色的一个课程，所有数据集，连同有关本书的其他信息，均可以从http://wooldridge.swcollege.com得到。

本书的一个独特的特色，是它有一个广泛的词汇表。当学生为应付考试或者阅读用到了计量经济学方法的经验研究论文时，表中的简短定义和描述将为他们提供一份有用的复习材料。

学生补充读物

附解答的学生学习指南（ISBN 0-538-85016-7）包含怎样阅读每一章的建议；部分习题和计算机作业的答案。

教师补充读物

附解答的教师手册（ISBN 0-538-85015-9）包含全部作业的答案；怎样讲授好每一章的教学心得。

感谢

从本书形成的最早期开始，许多人便已对它正式或非正式地评阅过。我要对他们表示感谢。在他们之中有：

Richard Agnello (University of Delaware)
Eli Berman (Boston University)
Christopher Cornwell (University of Georgia)
Edward Coulson (Pennsylvania State University)
William Even (Miami University of Ohio)
Adrian Fleissig (St. Louis University)
Heejoon Kang (Indiana University)
Neha Khanna (SUNY, Binghamton)
Kristin McCue (Texas A&M University)
John Mullahy (University of Wisconsin)
William Neilson (Texas A&M University)
David Neumark (Michigan State University)
Leslie Papke (Michigan State University)
Jeffrey Pliskin (Hamilton College)
Joseph Quinn (Boston College)
Nagesh Revankar (SUNY, Buffalo)
Louise Russell (Rutgers University)
Mark Showalter (Brigham Young University)
John Spitzer (SUNY, Brockport)
James Stock (Harvard University)
Christopher Taber (Northwestern University)
Larry Taylor (Lehigh University)
Pravin Trivedi (Indiana University)
Robert Trost (George Washington University)
Hiroki Tsurumi (Rutgers University)
Timothy Vogelsang (Cornell University)
Paul Wilson (University of Texas)
Keith Womar (University of Mississippi)

我还要感谢 Chirok Han，他阅读了这部书稿并改正了一些错误。自然，对全书内容以及尚存的错误，应由我来负最终的责任。

当我从 1986 年到 1991 年在麻省理工学院任教时，我便开始思考写一本入门计量经济学的教科书。现在我教的这门课和我当初第一次教的在许多重要方面都不相同。不论在密歇根州立大学或 MIT，学生的反应对本书的形成都起着重要作用。一些人甚至连他们自己都没有意识到，影响了本书的体裁和内容。这些人中有：加州大学伯克利分校的 Thomas Rothenberg，他教了我第一个计量经济学课程；我读研究生院时的导师 Robert Engle，加州大学圣迭戈分校的 Clive Grange，尤其是，Halbert White，他们教会我直觉和严谨是同等重要的；我在 MIT 时的同事 Jerry Hausman 和 Daniel McFadden，他们强调计量经济学也许对政策分析最为有用，以及 Arthur Goldberger，他的写作塑造了我对什么才是学计量经济学的现代方法的构思。

还有许多人，善意提供了数据集，在此无法一一列举。关于数据集来源，我在数据描述文件中作了特别的声明。

我怀念我的已故同事 Walter Adams，在我编写本科生教材的许多方面都有他的一份智慧，他坚信这项编写工作是有价值的。我作为密歇根州教学人员中的一员，Walter 使我感到更多的乐趣。

South-Western College 出版社的同仁们耐心支持着这一事业。我要感谢 Kenneth King 最先提出这项计划，但他现在离开了这个出版社。Jack Calhoun 从这项计划一开始就做了很好的技术性安排，然后 Thomas Sigel 在关键时刻来到编辑部，在推动本书写作并使之完成方面起到重要作用。Keri Witman 在本教科书的补充读物方面给予了有益的点拨。Peggy Buskey 做了出色的出品编辑者工作；Malvine Litten 及其在 LEAP 的同仁们细致并职业性地把一份难读的原稿变成现在这部教材。

最后，我以最愉快的心情感谢我的妻子 Leslie Papke，不仅是因为她的一贯支持和鼓励，还因为她不断为这部教材做出实质性的贡献。直接或间接地取自她在公共经济学中的实证研究的例子，一直在这部教材中闪光（包括在本书中的若干个数据集本来就是她的）。Leslie 还细心地阅读并改进了一些章节，同时她还要应付繁忙的家务和其他任务。

我把这部教材献给 Leslie 和我们的孩子 Edmund 和 Gwenyth，他们曾不休止地延缓了本书的完成。毕竟，是把时间分配给同 Neddie 和 Gwennie 一起玩，还是干我的计量经济学，我们没有发生争吵。

Jeffrey M. Wooldridge

作者简介

J.M. 伍德里奇（Jeffrey M. Wooldridge）　密歇根州立大学经济学教授，1991 年以来一直在该校任教。1986—1991 年伍德里奇博士曾任麻省理工学院经济学助教授。1982 年他以计算机科学与经济学为主攻方向而获加州大学伯克利分校艺术学士学位，并于 1986 年于加州大学圣迭戈分校获经济学博士学位。伍德里奇博士曾在国际知名期刊发表学术论文二二十篇，参与过多种书籍的篇章写作。他的获奖项目包括：Alfred P. 斯隆（Sloan）研究员基金，计量经济理论 Multa Scripsicl 奖金，应用计量经济学期刊的 R. 斯通（Stone）爵士奖，以及三次获 MIT 当年研究生班优秀教师奖。他还是计量经济学期刊（*Journal of Econometrics*）的资深会员。

伍德里奇博士一直是 *Journal of Businass and Economic Statistics* 的编委，*Economics Letters* 的计量经济学方面编委，并供职于 *Journal of Econometrics* 和 *Review of Economics and Statistics* 的编辑委员会。他还担任芝加哥 Arthur Andersen 和波士顿 Charles River Associates 两家公司的临时计量经济学顾问。

目录

第1章 计量经济学的性质与经济数据 …… 1

1.1 什么是计量经济学 …… 1

1.2 经验经济分析的步骤 …… 2

1.3 经济数据的结构 …… 5

横截面数据 …… 6

时间序列数据 …… 8

混合横截面数据 …… 9

综列或纵剖面数据 …… 10

对数据结构的评论 …… 12

1.4 计量经济分析中的因果关系和其他条件不变的概念 …… 12

小结 …… 16

关键术语 …… 17

第1篇 横截面数据的回归分析 …… 19

第2章 简单回归模型 …… 21

2.1 简单回归模型的定义 …… 21

2.2 普通最小二乘法的推导 …… 26

关于术语的注解 …… 33

2.3 OLS的操作技巧 …… 33

拟合值和残差 …… 34

OLS统计的代数性质 …… 35

拟合优度 …… 37

2.4 测量单位和函数形式 …… 38

改变测量单位对OLS统计量的影响 …… 38

在简单回归中加入非线性因素 …… 39

"线性"回归的含义 …… 42

2.5 OLS估计量的期望值和方差 …… 43

OLS的无偏性 …… 43

OLS估计量的方差 …… 48

误差方差的估计 …… 52

2.6 过原点回归 …… 54

小结 …… 55

关键术语 …… 56

习题 …… 56

计算机习题 …… 59

附录2A …… 60

第3章 多元回归分析：估计 …… 62

3.1 使用多元回归的动因 …… 63

含有两个自变量的模型 …… 63

含有 K 个自变量的模型 …… 65

3.2 普通最小二乘法的操作和解释 …… 67

如何得到OLS估计值 …… 67

对OLS回归方程的解释 …… 68

多元回归中"保持其他因素不变"的含义 …… 70

同时改变不止一个自变量 …… 71

OLS的拟合值和残差 …… 71

对多元回归"排除其他变量影响"的解释 …… 72

简单回归和多元回归估计值的比较 …… 73

拟合优度 …… 74

通过原点的回归 …… 77

3.3 OLS估计量的期望值 …… 77

在回归模型中包含了无关变量 …… 82

遗漏变量的偏误：简单情形 …… 83

遗漏变量的偏误：更一般的情形 …… 86

3.4 OLS估计量的方差 …… 87

OLS方差的成分：多重共线性 …… 89

误设模型中的方差 …… 92

估计 σ^2：OLS估计量的标准误 …… 93

3.5 OLS的有效性：高斯-马尔科夫定理 …… 95
小结 …… 96
关键术语 …… 97
习题 …… 97
计算机习题 …… 101
附录3A …… 103
第4章 多元回归分析：推断 …… 107
4.1 OLS估计量的抽样分布 …… 107
4.2 检验对单个总体参数的假设：t检验 …… 110
对单侧对立假设的检验 …… 112
双侧对立假设 …… 117
检验β_j的其他假设 …… 119
计算t检验的p值 …… 122
对经典假设检验用语的提醒 …… 124
经济或实际显著性与统计显著性 …… 124
4.3 置信区间 …… 127
4.4 检验关于参数的一个线性组合的假设 …… 129
4.5 对多个线性约束的检验：F检验 …… 132
对排除性约束的检验 …… 132
F统计量和t统计量之间的关系 …… 138
F统计量的R-平方型 …… 139
计算F检验的p值 …… 140
回归整体显著性的F统计量 …… 141
检验一般的线性约束 …… 142
4.6 报告回归结果 …… 143
小结 …… 145
关键术语 …… 146
习题 …… 147
计算机习题 …… 152
第5章 多元回归分析：OLS的渐近性 …… 154
5.1 一致性 …… 155
推导OLS的不一致性 …… 157
5.2 渐近正态和大样本推断 …… 159
其他大样本检验：拉格朗日乘数统计量 …… 163
5.3 OLS的渐近有效性 …… 165
小结 …… 167
关键术语 …… 167
习题 …… 168
计算机习题 …… 168

附录 5A …… 169

第 6 章 多元回归分析：其他问题 …… 171

6.1 数据的测度单位对 OLS 统计量的影响 …… 171

β 系数 …… 174

6.2 对函数形式的进一步讨论 …… 176

对使用对数函数形式的进一步讨论 …… 176

含二次式的模型 …… 178

含有交互作用项的模型 …… 182

6.3 拟合优度和回归元选择的进一步探讨 …… 184

调整 R-平方 …… 184

利用调整 R-平方在两个非嵌套模型中进行选择 …… 186

回归分析中控制了过多的因素 …… 188

增加回归元以减少误差方差 …… 189

6.4 预测和残差分析 …… 190

预测的置信区间 …… 190

残差分析 …… 193

当因变量为log(y)时对 y 的预测 …… 194

小结 …… 197

关键术语 …… 198

习题 …… 198

计算机习题 …… 200

第 7 章 含有定性信息的多元回归分析：二值（或虚拟）变量 …… 203

7.1 对定性信息的描述 …… 204

7.2 只有一个虚拟自变量 …… 205

当因变量为 log(y)时，对虚拟解释变量系数的解释 …… 210

7.3 使用多个虚拟变量 …… 211

通过虚拟变量来包含序数信息 …… 213

7.4 涉及虚拟变量的交互作用 …… 216

虚拟变量之间的交互作用 …… 216

容许出现不同的斜率 …… 217

检验不同组之间回归函数上的差别 …… 221

7.5 二值因变量：线性概率模型 …… 224

7.6 对政策分析和项目评价的进一步讨论 …… 229

小结 …… 231

关键术语 …… 232

习题 …… 232

计算机习题 …… 235

第 8 章 异方差性 …… 238

8.1 异方差性对 OLS 所造成的影响 …… 238

8.2 OLS 估计后异方差—稳健性推断 …… 239

计算异方差—稳健的 LM 检验 …… 243

8.3 对异方差性的检验 …… 245

异方差性的 White 检验 …… 248

8.4 加权最小二乘估计 …… 250

除一个常数倍数外异方差是已知的 …… 251

必须估计异方差函数：可行 GLS …… 256

8.5 再议线性概率模型 …… 260

小结 …… 262

关键术语 …… 263

习题 …… 263

计算机习题 …… 264

第 9 章 模型设定和数据问题的深入探讨 …… 267

9.1 函数形式误设 …… 268

对函数形式误设问题的一般检验：RESET …… 270

对非嵌套模型的检验 …… 272

9.2 对观测不到的解释变量使用代理变量 …… 273

用滞后因变量作为代理变量 …… 277

9.3 有测量误差的 OLS 性质 …… 279

因变量中的测量误差 …… 280

解释变量中的测量误差 …… 282

9.4 数据缺失、非随机样本和异常观测 …… 285

数据缺失 …… 286

非随机样本 …… 286

异常观测 …… 288

小结 …… 292

关键术语 …… 293

习题 …… 293

计算机习题 …… 295

第 2 篇 时间序列数据的回归分析 …… 297

第 10 章 时间序列数据的基本回归分析 …… 299

10.1 时间序列数据的性质 …… 299

10.2 时间序列回归模型的例子 …… 301

静态模型 …… 301

有限分布滞后模型 …… 301

标注时间的惯例 …… 304

10.3 经典假设下 OLS 的有限样本性质 …… 304

OLS 的无偏性 …… 304

OLS 估计量的方差和高斯-马尔科夫定理 …… 308

经典线性模型假定下的推断 …… 310

10.4 函数形式、虚拟变量和指数 …… 311
10.5 趋势和季节性 …… 318
描述有趋势的时间序列 …… 318
在回归分析中使用趋势变量 …… 321
对有时间趋势的回归做除趋势变换 …… 323
因变量有趋势时 R-平方的计算 …… 325
季节性 …… 326
小结 …… 328
关键术语 …… 329
习题 …… 329
计算机习题 …… 330

第 11 章 用时间序列数据计算 OLS 的其他问题 …… 333

11.1 平稳性和弱相依时间序列 …… 334
平稳和非平稳时间序列 …… 334
弱相依时间序列 …… 335
11.2 OLS 的渐近性质 …… 338
11.3 使用高度持久时间序列做回归分析 …… 344
高度持久时间序列 …… 344
高度持久时间序列的变换 …… 348
判断时间序列是否是I(1) …… 349
11.4 动态完整模型和无序列相关 …… 351
11.5 时间序列模型的同方差假定 …… 354
小结 …… 354
关键术语 …… 355
习题 …… 355
计算机习题 …… 358

第 12 章 时间序列回归中的序列相关和异方差 …… 361

12.1 有序列相关误差的 OLS 性质 …… 362
无偏性和一致性 …… 362
效率和推断 …… 362
出现滞后因变量时的序列相关 …… 363
12.2 序列相关的检验 …… 365
回归元为严格外生时对AR(1)序列相关的 t 检验 …… 365
经典假定条件下的德宾-沃森统计量 …… 367
回归元不是严格外生时AR(1)序列相关的检验 …… 369
更高阶序列相关的检验 …… 370
12.3 对严格外生回归元的序列相关的校正 …… 372
在AR(1)模型中求最优线性无偏估计量 …… 372
有AR(1)误差的可行 GLS 估计 …… 374
OLS 和 FGLS 的比较 …… 376

更高阶序列相关的校正 …… 378

12.4 差分和序列相关 …… 379

12.5 在OLS后的序列相关—稳健推断 …… 380

12.6 时间序列回归中的异方差性 …… 383

异方差-稳健统计量 …… 384

异方差的检验 …… 384

自回归条件异方差 …… 385

回归模型中的异方差和序列相关 …… 387

小结 …… 388

关键术语 …… 389

习题 …… 389

计算机习题 …… 390

第3篇 高深专题讨论 …… 393

第13章 跨时横截面的混合，简单综列数据方法 …… 395

13.1 跨时独立横截面的混合 …… 396

对跨越时间的结构性变化做邹至庄检验 …… 400

13.2 利用混合横截面做政策分析 …… 401

13.3 两时期综列数据分析 …… 406

综列数据的编排 …… 412

13.4 用两期综列数据做政策分析 …… 412

13.5 多于两期的差分法 …… 415

小结 …… 420

关键术语 …… 420

习题 …… 421

计算机习题 …… 422

附录13A …… 425

第14章 高深的综列数据方法 …… 427

14.1 固定效应估计法 …… 427

虚拟变量回归 …… 431

是固定效应（FE）还是一阶差分（FD）? …… 433

非平衡综列数据的固定效应法 …… 434

14.2 随机效应模型 …… 435

随机效应还是固定效应? …… 438

14.3 把综列数据方法用于其他数据结构 …… 439

小结 …… 441

关键术语 …… 441

习题 …… 442

计算机习题 …… 443

附录14A …… 445

第 15 章　工具变量估计与两阶段最小二乘法 …… 447
15.1　动机：简单回归模型中的遗漏变量 …… 448
用 IV 估计量做统计推断 …… 451
低劣的工具变量条件下 IV 的性质 …… 456
IV 估计后计算 R^2 …… 457
15.2　多元回归模型的 IV 估计 …… 458
15.3　两阶段最小二乘 …… 461
单一内生解释变量 …… 461
多重共线性与 2SLS …… 464
多个内生解释变量 …… 465
2SLS 估计后对多个假设的检验 …… 466
15.4　含误差的变量问题的 IV 解 …… 466
15.5　内生性检验与检验过度识别约束 …… 468
内生性检验 …… 468
检验过度识别约束 …… 470
15.6　异方差性条件下的 2SLS …… 471
15.7　2SLS 应用于时间序列方程 …… 472
15.8　2SLS 应用于混合横截面和综列数据 …… 474
小结 …… 476
关键术语 …… 477
习题 …… 477
计算机习题 …… 480
附录 15A …… 483
第 16 章　联立方程模型 …… 485
16.1　联立方程模型的性质 …… 486
16.2　OLS 中的联立性偏误 …… 489
16.3　结构方程的识别和估计 …… 491
两方程联立模型中的识别 …… 491
使用 2SLS 的估计 …… 496
16.4　多于两个方程的联立方程组 …… 497
三个或更多方程的系统中的识别问题 …… 498
估计 …… 499
16.5　利用时间序列的联立方程模型 …… 499
16.6　利用综列数据的联立方程模型 …… 503
小结 …… 505
关键术语 …… 506
习题 …… 506
计算机习题 …… 508
第 17 章　限值因变量模型和样本选择纠正 …… 511
17.1　二值响应的 logit 和 probit 模型 …… 512

设定 logit 和 probit 模型 …… 512
logit 和 probit 模型的最大似然估计 …… 515
多重假设的检验 …… 516
解释 logit 和 probit 模型的估计值 …… 517
17.2 Tobit 模型 …… 521
对 Tobit 估计值的解释 …… 522
Tobit 模型中的设定问题 …… 526
17.3 泊松回归模型 …… 527
17.4 截取和断尾回归模型 …… 531
截取回归模型 …… 532
断尾回归模型 …… 535
17.5 样本选择纠正 …… 537
OLS 什么时候对选择的样本是一致的？ …… 538
偶然断尾 …… 539
小结 …… 543
关键术语 …… 544
习题 …… 545
计算机习题 …… 546
附录 17A …… 548
第 18 章 时间序列的深入讨论 …… 550
18.1 无限分布滞后模型 …… 551
几何（或考依克）分布滞后 …… 553
有理分布滞后模型 …… 555
18.2 单位根的检验 …… 557
18.3 谬误回归 …… 563
18.4 协积和误差纠正机制 …… 565
协积 …… 565
误差纠正模型 …… 570
18.5 预测 …… 572
用于预测的各种回归模型 …… 573
超前一步预测 …… 574
超前一步预测的比较 …… 578
超前多步预测 …… 579
有趋势、季节性和自积过程的预测 …… 582
小结 …… 587
关键术语 …… 588
习题 …… 589
计算机习题 …… 591
第 19 章 一个经验项目的实施 …… 594
19.1 问题的提出 …… 594

19.2 文献回顾 …… 597
19.3 数据的收集 …… 597
确定适当的数据集 …… 597
输入并储存你的数据 …… 599
检查、清理、总结你的数据 …… 600
19.4 计量经济学分析 …… 602
19.5 经验论文的写作 …… 604
引言 …… 605
概念（或理论）框架 …… 605
计量经济学模型和估计方法 …… 606
数据 …… 608
结果 …… 609
结论 …… 610
风格提示 …… 610
小结 …… 613
关键术语 …… 613
样本经验项目 …… 613
期刊列表 …… 618
数据资源 …… 619
附录A 基本数学工具 …… 621
A.1 总和运算子与描述统计量 …… 621
A.2 线性函数的性质 …… 624
A.3 比例与百分数 …… 626
A.4 若干特殊函数及其性质 …… 628
二次函数 …… 628
自然对数 …… 630
指数函数 …… 633
A.5 微分学 …… 634
小结 …… 637
关键术语 …… 637
习题 …… 637
附录B 概率论基本知识 …… 640
B.1 随机变量及其概率分布 …… 641
离散随机变量 …… 641
连续随机变量 …… 643
B.2 联合分布、条件分布与独立性 …… 645
联合分布与独立性 …… 645
条件分布 …… 647
B.3 概率分布的特征 …… 648
集中趋势的一种度量：期望值 …… 648

期望值的性质 …… 649
集中趋势的另一种度量：中位数 …… 651
变异性的度量：方差与标准差 …… 652
方差 …… 652
标准差 …… 654
标准化一个随机变量 …… 654
B.4 联合与条件分布的特征 …… 655
关联度：协方差与相关 …… 655
协方差 …… 655
相关系数 …… 656
随机变量之和的方差 …… 657
条件期望 …… 659
条件期望的性质 …… 660
条件方差 …… 662
B.5 正态及其有关分布 …… 663
正态分布 …… 663
标准正态分布 …… 664
正态分布的其他性质 …… 666
χ 平方分布 …… 666
t 分布 …… 667
F 分布 …… 668
小结 …… 669
关键术语 …… 669
习题 …… 670
附录C 数理统计基础 …… 672
C.1 总体、参数与随机抽样 …… 672
抽样 …… 673
C.2 估计量的有限样本性质 …… 674
估计量与估计值 …… 674
无偏性 …… 675
估计量的抽样方差 …… 677
有效性 …… 679
C.3 估计量的渐近或大样本性质 …… 680
一致性 …… 681
渐近正态性 …… 683
C.4 参数估计的一般方法 …… 685
矩法 …… 685
最大似然法 …… 686
最小二乘法 …… 687

C.5 区间估计与置信区间 …… 687
区间估计的性质 …… 687
正态分布总体均值的置信区间 …… 689
95%置信区间的一个简单的经验法则 …… 692
非正态总体的渐近置信区间 …… 693
C.6 假设检验 …… 694
假设检验的基本知识 …… 694
检验关于正态总体均值的假设 …… 696
非正态总体的渐近检验 …… 699
p 值的计算和使用 …… 700
置信区间与假设检验的关系 …… 703
实际显著性与统计显著性的对比 …… 704
C.7 关于符号的注释 …… 705
小结 …… 706
关键术语 …… 706
习题 …… 707
附录D 矩阵代数概述 …… 713
D.1 基本定义 …… 713
D.2 矩阵运算 …… 714
矩阵加法 …… 714
数乘 …… 715
矩阵乘法 …… 715
转置 …… 717
分块矩阵的乘法 …… 717
迹 …… 718
逆 …… 718
D.3 线性独立与矩阵的秩 …… 719
D.4 二次型与正定矩阵 …… 720
D.5 幂等矩阵 …… 720
D.6 线性形式和二次型的微分 …… 721
D.7 随机向量的矩和分布 …… 721
期望值 …… 721
方差—协方差矩阵 …… 722
多元正态分布 …… 722
χ 平方分布 …… 723
t 分布 …… 723
F 分布 …… 723
小结 …… 724
关键术语 …… 724
习题 …… 724

附录E　矩阵形式的线性回归模型 …… 726
E.1　模型与普通最小二乘估计 …… 726
E.2　OLS的有限样本性质 …… 729
E.3　统计推断 …… 732
小结 …… 734
关键术语 …… 734
习题 …… 734
附录F　各章习题解答 …… 736
附录G　统计学用表 …… 751
参考文献 …… 759
术语表 …… 767
索引 …… 781
译后记 …… 827

第 1 章　计量经济学的性质与经济数据

第 1 章讨论的是计量经济学的研究领域，并提出在应用计量经济方法过程中所遇到的一般问题。1.3 节考察了商业、经济学和其他社会科学中所使用的数据集的种类。1.4 节对社会科学中的因果性推断的困难进行了直观讨论。

1.1　什么是计量经济学

1　设想州政府雇用了你，要求你对公共机构的在职培训项目的效果进行评价。假设这个项目是培训工人在生产过程中使用计算机的各种方法。为期 20 周的培训都是在工人的非工作时间进行，任何一个按小时计酬的生产工人都可以自愿参加全部或部分培训。你要决定培训项目对每个工人随后的小时工资有何影响。

现在假设你为一家投资银行工作，并准备研究涉及美国短期国库券的各种不同投资战略的回报，看看它们是否与经济理论的含义相一致。

回答这种问题的任务初看起来让人胆怯。就此看来，你对需要搜集的数据类型只有模糊的认识。在学完这本计量经济学入门教程后，你应该知道，

如何用计量模型去规范地评价一个在职培训项目，或检验一个简单的计量经济理论。

计量经济学对经济关系的估计、对经济理论的检验以及对政府和商业政策的评价和实施，都取决于统计方法的进步。计量经济学最常见的应用就是对诸如利率、通货膨胀率和国内生产总值等重要宏观经济变量的预测。尽管对经济指标的预测随处可见而又广为流传，但计量经济方法也可用于那些与宏观经济预测无关的经济领域。比如，我们可以研究在政治竞争中支出对投票结果的影响，还可以考虑学校支出对教育领域学生表现的影响。此外，我们还将了解如何使用计量方法来预测经济时间序列。

计量经济学已从数理统计分离出来并演化成一门独立学科，因为前者主要考虑在搜集和分析非实验经济数据时的固有问题。**非实验数据**（nonexperimental data）并非从对个人、企业或经济系统中的某些部分的控制实验而得来。[非实验数据有时被称为**观测数据**（observational data），以强调研究者只 2 是被动的数据搜集者这一事实。] 自然科学中的**实验数据**（experimental data）通常是在实验环境中获得的，但在社会科学中要得到这些实验数据则困难得多。虽然也可以设计一些社会实验，但通常都行不通，它代价高昂而使人望而却步，或者解决经济问题所需要实施的各种控制实验在道德上使人极为反感。在 1.4 节我们给出一些特殊的例子，来说明实验数据与非实验数据之间的差别。

当然，计量经济学家只要有可能就会借用数理统计学家的一些方法。多元回归分析方法虽然在上述两个领域都是主要支柱，但其着眼点和解释可以极为不同。此外，经济学家已想出许多新方法，以处理经济数据的复杂性和检验经济理论所预测的结果。

1.2 经验经济分析的步骤

计量经济方法几乎在应用经济学的每一个分支中都相当重要。要么在我们有一个经济理论需要检验的时候，要么在我们的脑海中有一种关系，而在这一关系对商业决策或政策分析又相当重要的时候，便开始使用计量方法。**经验分析**（empirical analysis）* 就是利用数据来检验某个理论或估计某种关系。

应该如何构建一个经验经济分析呢？虽然看上去十分显然，但仍值得强调，任何一个经验分析的第一步都是对所关心问题的详细阐述。问题可能是要处理对一个经济理论某特定方面的检验，或者涉及对政府政策效果的检验。原则上，计量经济方法可用来回答诸多方面的问题。

在某些情形下，特别是涉及对经济理论的检验时，就要构造一个规范的

---

* 国内常称实证分析，但为了区别于 positive analysis，故译为经验分析。——译者注

经济模型（economic model)。经济模型总是由描述各种关系的数理方程构成。经济学家因建造模型来描述大量人类行为而声名远扬。例如在中级微观经济学中，个人在预算约束下的消费决策便可由一些数理模型来描述。这些模型背后的基本前提是**效用最大化**。个人选择能在资源约束条件下最大化其福利的假定，为我们创造一些简便的经济模型和作出一些明确的预测提供了强有力的框架。在消费决策的背景下，效用最大化能导出一系列**需求方程**。在每个需求方程中，每种商品的需求量取决于该商品的价格、其替代品和互补品的价格、消费者的收入和影响消费者个人喜恶的一些个人特征。这些方程便形成对消费需求进行计量分析的基础。

经济学家使用诸如效用最大化框架之类的基本分析工具，来解释那些初看起来具有非经济性质的行为。一个经典的例子就是贝克尔（Becker，1968）针对犯罪行为而做的经济模型。

---

3 **例 1.1　犯罪的经济模型**

在一篇开创性的论文中，诺贝尔经济学奖得主加里·贝克尔（Gary Becker）系统地阐述了一个效用最大化框架，用以描述个人对犯罪行为的选择。虽然每一特定的犯罪都有明显的经济回报，但大多数犯罪行为也有其成本。犯罪的机会成本使罪犯不能参加诸如合法就业之类的其他活动。此外，还存在与罪犯可能被抓住相联系的成本，以及罪犯被抓后，如果被证明有罪，与监禁相关的成本。从贝克尔的视角来看，决定进行非法活动的决策是资源配置的方式之一，并是在充分考虑了各种可选择行为的成本和收益后决定的。

在一般化的假定之下，我们便能推导出一个方程，把花在犯罪活动上的时间描述成各种影响因素的一个函数。我们可以把这个方程表示为

$$y=f(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \tag{1.1}$$

式中：y——花在犯罪活动上的小时数

x_1——从事犯罪活动每小时的“工资”

x_2——合法就业的小时工资

x_3——犯罪或就业之外的收入

x_4——犯罪被抓住的概率

x_5——犯罪被抓后，被证明有罪的概率

x_6——被证明有罪后预期的宣判

x_7——年龄

虽然还有其他因素通常会影响个人参与犯罪的决策，但上述因素从规范的经济分析来看可能具有代表性。如经济理论的惯常做法那样，我们未对式(1.1)中的函数$f(\cdot)$进行任何设定。这个函数取决于一个很少有人知道的潜在效用函数。尽管如此，我们还是可以用经济理论（或反思）来预测每个变量对犯罪活动可能具有的影响。这正是对个人犯罪行为进行计量经济分析的基础。

虽然规范的经济建模有时是经验分析的起点，但更普遍的情况是，使用经济理论时不是那么规范，甚至完全是依赖直觉。你可能也同意，方程（1.1）中出现的犯罪行为的决定因素从常识来看也是合情合理的；我们也许能直接得到这个方程，而不需要从效用最大化开始把它推导出来。尽管在有些情况下，规范的推导能提供直觉看不到的洞见，但这种观点也有其优点。

下面这个例子中的方程，就是从多少有些不甚规范的推理中得到的。

---

4 **例 1.2　工作培训与工人的生产力**

现在考虑在 1.1 节之初提出的问题。一位劳动经济学家想考察工作培训对工人生产力的影响。在此情形下，几乎不需要什么规范的经济理论。基本的经济常识就足以使我们认识到，所受教育、工作经历和培训等因素会影响工人的生产力。此外，经济学家还清楚地知道，工人的工资与其生产力相称。这种简单的理由就使我们得到如下模型：

$$wage = f(educ,\ exper,\ training) \tag{1.2}$$

式中，*wage* 为小时工资；*educ* 为接受正规教育的年限；*exper* 为工作年数；*training* 为花在工作培训上的周数。此外，虽然也有其他因素通常会影响工资率，但式（1.2）刻画了这个问题的本质。

在设定一个经济模型之后，需要把它变成所谓的**计量模型**（econometric model）。既然我们在全书都要讨论计量模型，那么先了解一下计量模型和经济模型有何关系就很重要。以方程（1.1）为例，与经济分析不同，在进行计量经济分析之前，我们必须明确函数 $f(\cdot)$ 的形式。与方程（1.1）相关的第二个问题是，对不能合理地观测到的变量该如何处理。比如，考虑一个人在进行犯罪活动时的工资。原则上，这个工资是清楚界定的，但对一个特定的人来说，这个工资是很难观测到的，甚至是不可能观测到的。虽然对某给定个人，诸如其被抓住的概率之类的变量也不能切实得到，但至少我们能找到相关的逮捕统计量，从而推导出一个能近似被抓住概率的变量。还有其他影响犯罪行为的因素，不要说观测，甚至连列出来都做不到，但我们多少都要对它们作出解释。

通过设定一个特定的计量经济模型，我们就解决了经济模型中内在的模棱两可性：

$$\begin{aligned} crime = {} & \beta_0 + \beta_1 wage_m + \beta_2 othinc + \beta_3 freqarr + \beta_4 freqconv \\ & + \beta_5 avgsen + \beta_6 age + u \end{aligned} \tag{1.3}$$

式中，*crime* 为参与犯罪活动频率的某种度量；$wage_m$ 为在合法就业中所得到的工资；*othinc* 为通过其他途径得到的收入（如资产、继承等）；*freqarr* 为以前违法被抓住的概率（用来近似被捕概率）；*freqconv* 为被证明有罪的概率；*avgsen* 为被证明有罪后判刑的平均时间长度。对这些变量的选择，既以经济理论为依据，又考虑到了数据。u 这一项则包括了不可观测的因素，诸如从事犯罪活动的工资、道德特征、家庭背景等，以及在度量犯罪活动和

被捕概率等变量时的误差。虽然我们也可以在模型中加入家庭背景变量，如兄弟姐妹的个数、父母所受教育等，但我们仍不能完全消除 u。实际上，对这个**误差项**（error term）或**扰动项**（disturbance term）的处理可能是任何计量分析中最重要的内容。

5 常数 β_0、β_1，…，β_6 都是这个计量模型的参数，它们描述了此模型中用来决定犯罪的因素和犯罪之间关系的方向和强度。

对例 1.2 来说，一个完整的计量经济模型可能是

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u \tag{1.4}$$

其中 u 这一项包含的因素有天生能力、教育的质量、家庭背景等，以及能影响一个人工资的无数其他因素。如果专门考虑工作培训的影响，那 β_3 就是我们所关注的参数。

在多数情况下，计量经济分析是从对一个计量经济模型的设定开始的，而没有考虑模型构造的细节。我们一般遵循这一思路，主要原因是，对犯罪这种经济模型进行仔细的推导，不仅消耗的时间过长，而且会把我们带到经济理论的某个特定而通常又极为困难的领域。在我们的例子中，经济上的逻辑如要起作用，那我们就要把其背后的任何一种经济理论都放到对计量模型的设定中。而在犯罪一例的经济模型中，我们将从像式（1.3）那样的一个计量模型出发，并以经济逻辑和常识作为选择变量的向导。尽管这一方法有失经济分析之丰富，但它总是被仔细的研究者普遍而又有效地应用着。

一旦设定了一个像式（1.3）或式（1.4）那样的计量模型，我们所关心的各种**假设**（hypotheses）便可用未知参数来表述。比如，在方程（1.3）中，我们可以假设合法就业的工资 $wage_m$ 对犯罪行为没有影响。在这个特定的计量模型背景下，这个假设就等价于 $\beta_1 = 0$。

按定义，一项经验分析总需要数据。在搜集到相关变量的数据之后，便用计量方法来估计计量模型中的参数，并规范地检验所关心的假设。在某些情况下，计量模型还用于对理论的检验或对政策影响的研究。

由于数据搜集在经验工作中如此重要，所以 1.3 节将专门介绍我们可能会遇到的数据类型。

1.3 经济数据的结构

经济数据的类型各式各样。尽管有些计量方法可以在对许多不同的数据集不做任何修改的情况下使用，但仍有必要对某些数据集的特殊性质进行阐释并加以利用。接下来我们就描述在应用研究中可能会遇到的几种重要的数据结构。

横截面数据

所谓**横截面数据集**（cross-sectional data set），就是在给定时点对个人、家庭、企业、城市、州、国家或一系列其他单位采集的样本所构成的数据 6 集。有时，所有单位的数据并非完全对应于同一时间段。例如，几个家庭可能在一年中的不同星期被调查。在一个纯粹的横截面分析中，我们应该忽略数据搜集中细小的时间差别。如果一系列家庭都是在同一年度的不同星期被调查的，那我们仍视之为横截面数据集。

横截面数据的一个重要特征是，我们通常可以假定，它们是从样本背后的总体中通过**随机抽样**（random sampling）而得到的。例如，如果我们通过随机地从工人总体中抽取 500 人，并得到其有关工资、受教育程度、工作经历和其他特征方面的信息，就得到所有工人构成的总体的一个随机样本。随机抽样是初级统计学教程中所讲授的抽样方案，而且它使得对横截面数据的分析大为简化。附录 C 对随机抽样进行了复习。

有时，以随机抽样作为对横截面数据的一个假定并不适当。例如，假设我们对研究影响家庭财富积累的因素感兴趣，虽然可以调查家庭的一个随机样本，但有些家庭可能拒绝报告其财富。比方说，如果越是富裕的家庭就越不愿意暴露其财富，那么由此得到的财富样本就不是由所有家庭构成的总体的一个随机样本。这是对样本选择问题的一个解释，我们还将在第 17 章专门讨论这个高级专题。

当我们抽取的样本（特别是地理上的样本）相对总体而言太大时，可能会导致另一种偏离随机抽样的情况。这种情形中潜在的问题是，总体不够大，所以不能合理地假定观测值是独立抽取的。例如，如果我们想用工资率、能源价格、公司和财产税、所提供的服务、工人的质量及其他有关州的特征来解释各州间新的商业活动，那么，邻近州之间的商业活动不可能相互独立。虽然事实表明我们所讨论的计量方法在这种情形下确实能起作用，但有时这些方法还需要提炼。多数情况下，我们都放弃分析这种情形所引起的错综复杂的关系，而在随机抽样的框架下处理这些问题，尽管有时这样做在技术上是不正确的，但我们只好如此。

横截面数据被广泛地应用于经济学和其他社会科学领域之中。在经济学中，横截面数据分析与应用经济领域有密切联系，如劳动经济学、州和地方公共财政学、产业组织理论、城市经济学、人口和健康经济学。对检验微观经济假设和评价经济政策而言，在一给定时点上，有关个人、家庭、企业和城市的数据都是至关重要的。

用于计量经济分析的横截面数据可以在计算机上表示和存储。表 1.1 以缩略的形式给出了 1976 年 526 个工人的横截面数据集。（这是文件 WAGE1.RAW 中数据的一个子集。）变量包括工资 *wage*（美元/小时）、受教育年数 *educ*、潜在劳动经历的年数 *exper*、性别 *female*（1 表示女性、0 表示男性）

和婚姻状况 *married*（1表示已婚，0表示未婚）。后面两个变量在本质上是二值（0—1）变量，并表明个人的定性特征（这个人是不是女性；这个人是不是结了婚）。在第7章及以后，我们会详谈二值变量。

表1.1中的变量 *obsno* 是赋予样本中每个人的观测序号，与其他变量不同，它不是个人特征。所有的计量经济学和统计学软件包，对每个数据单位都会指定一个观测序号。凭直觉，对表1.1中的数据而言，哪个被标为观测1，哪个被标为观测2等等，应该是没有多大问题的。数据排序不影响计量分析这一事实，是随机抽样而得到横截面数据集的一个重要特征。

7 **表1.1　有关工资和其他个人特征的横截面数据集**

obsno	*wage*	*educ*	*exper*	*female*	*married*
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

在横截面数据集中，不同的变量有时对应于不同的时期。例如，为了决定政府政策对长期经济增长的影响，经济学家研究了一定时期（如1960—1985年）真实人均国内生产总值（GDP）的增长率与部分由1960年的政府政策（政府消费占GDP的百分比和成人中受过中等教育的百分比）所决定的变量之间的关系。这样一个数据集可由表1.2表示，这些正是德·朗和萨默斯（De Long and Summers, 1991）对各国经济增长率的研究中所用到的数据集的一部分。

8 **表1.2　经济增长率和国家特征方面的数据集**

obsno	*country*	*gpcrgdp*	*Govcons*60	*Second*60
1	阿根廷	0.89	9	32
2	澳大利亚	3.32	16	50
3	比利时	2.56	13	69
4	玻利维亚	1.24	18	12
⋮	⋮	⋮	⋮	⋮
61	津巴布韦	2.30	17	6

表中，变量 *gpcrgdp* 为 1960－1985 年间真实人均 GDP 的平均增长率。尽管 *govcons*60（政府消费占 GDP 的百分比）和 *second*60（成人中受过中等教育的百分比）都对应于 1960 年，而 *gpcrgdp* 则是 1960—1985 年的平均增长率，但把这些信息看成横截面数据集并不会导致任何特别的问题。虽然观测值以国家名称的字母排序，但这种排序对以后的分析没有任何影响。

时间序列数据

时间序列数据集（time series data set）是由一个或几个变量不同时间的观测值所构成。时间序列数据方面的例子包括股票价格、货币供给、消费者价格指数、国内生产总值、年度谋杀率和汽车销售数量。由于过去的事件可以影响到未来的事件，而且行为滞后在社会科学中又相当普遍，所以时间是时间序列数据集中的一个重要维度。与横截面数据的排序不同，时间序列对观测值按时间先后排序也传递了潜在的重要信息。

时间序列数据有一个关键的特征，使得对它的分析比对横截面数据的分析更为困难。这个特征就是如下事实，即很少（即使能够）假设经济数据的观测独立于时间。多数经济上和其他的时间序列都与其近期历史相关（通常是高度相关）。比如，由于 GDP 的趋势从这个季度到下个季度保持着相当的稳定性，所以对上一季度 GDP 的一些了解，会告诉我们本季度 GDP 的可能范围。虽然多数计量程序既能用于横截面数据，又能用于时间序列数据，但在认为标准的计量方法能适用之前，要设定时间序列数据的计量模型，还有更多工作要做。此外，为了解释和利用经济时间序列的相互依赖性，并解决某些经济变量常表现出清晰的时间趋势等问题，对标准的计量方法还要进行改进和润色。

需要特别注意的是时间序列数据的另一个特征，即数据搜集中的**数据频**
9 **率**（data frequency），最常见的频率是每天、每周、每月、每个季度和每年。股票价格以天为区间进行记录（星期六和星期日除外）。美国的货币供给是逐月报告的，许多宏观经济序列都是按月列出，如通货膨胀率和就业率。其他宏观序列报告就没那么频繁，比如国内生产总值就是典型地每 3 个月（一个季度）报告一次。其他的时间序列，如美国各州的婴儿死亡率等，则只有年度数据可供使用。

许多按周、按月、按季度报告的经济时间序列都表现出很强的季节类型，这正是时间序列分析中的一个重要因素。比如，各个月份的新住房动工数据之不同，仅仅是因为气温条件的变化所致。在第 10 章中，我们将会了解如何处理这种季节性时间序列。

表 1.3 包含的时间序列数据集，来自卡斯蒂罗-弗里曼和弗里曼（Castillo Freeman and Freeman，1992）研究波多黎各的最低工资条例影响的一篇论文。此数据集中最早的那一年就是第一次观测，最后的那一年则是最后一次观测。在用计量方法分析时间序列数据时，数据应该按时间顺序排列。

表 1.3 波多黎各的最低工资、失业及相关数据

obsno	*year*	*avgmin*	*avgcov*	*unemp*	*gnp*
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1 015.9
⋮	⋮	⋮	⋮	⋮	⋮
37	1986	3.35	58.1	18.9	4 281.6
38	1987	3.35	58.2	16.8	4 496.7

表中，变量 *avgmin* 为当年的最低工资；*avgcov* 为最低工资的覆盖率（最低
10 工资法所包括的工人占工人总数的百分比）；*unemp* 为失业率；而 *gnp* 则是
国民生产总值。我们将在以后研究最低工资对就业之影响的一个时间序列分
析中使用这些数据。

混合横截面数据

有些数据既有横截面数据的特点又有时间序列的特点。例如，假设对美国的家庭进行了两次横截面数据的调查，一次在 1985 年，一次在 1990 年。在 1985 年，对家庭的一个随机样本调查了工资、储蓄、家庭大小等变量。到了 1990 年，用同样的调查问题又对家庭的一个新随机样本进行调查。为了扩大我们的样本容量，可以将这两年的数据合并成一个**混合横截面数据**(pooled cross section)。由于在每一年都是进行随机抽样，所以同一个家庭在两年的样本中都出现则纯属偶然。（与美国家庭的数量相比，样本容量通常都很小。）这一点使混合横截面数据有别于综列数据集。

把不同年份的横截面数据混合起来，通常是分析一项新政府政策之影响的有效方法。其思想是，搜集一个重要的政策变化之前和之后的数据。举例而言，考虑在 1993 年和 1995 年都对住房价格搜集了数据集，而在 1994 年则下调了财产税。假设我们在 1993 年有 250 个房子的数据，1995 年有 270 个房子的数据。表 1.4 给出了存储这种数据的一种方式。

观测 1～250 对应于 1993 年出售的房子，而观测 251～520 则对应于 1995 年出售的房子。尽管最终表明我们如何存放数据的顺序并不重要，但能区分出每一观测值的年份一般地说却十分重要，这就是为什么我们把年份作为一个独立的变量收入数据集。

对混合横截面数据的分析与对标准横截面数据的分析十分相似，不同之处在于，前者通常要对变量在不同时间的现实差异作出解释。实际上，除了

能扩大样本容量之外，混合横截面分析通常是为了让我们看出一个关键的关系如何随时间而变化。

表 1.4　　混合横截面数据：两年的住房价格

11

obsno	*year*	*hprice*	*proptax*	*sqrft*	*bdrms*	*bthrms*
1	1993	85 500	42	1 600	3	2.0
2	1993	67 300	36	1 440	3	2.5
3	1993	134 000	38	2 000	4	2.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮
250	1993	243 600	41	2 600	4	3.0
251	1995	65 000	16	1 250	2	1.0
252	1995	182 400	20	2 200	4	2.0
253	1995	97 500	15	1 540	3	2.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
520	1995	57 200	16	1 100	2	1.5

综列或纵剖面数据

综列数据（panel data）（或纵剖面数据）集由横截面数据集中每个数据的一个时间序列组成。比方说我们对一系列个人的工资、受教育情况和就业史跟踪了10年，或者我们对一系列企业诸如投资和财务数据等搜集了5年的信息。有些综列数据也可以以地理上的单位来搜集。比如，在1980年、1985年和1990年对美国同一组县搜集其人口迁移、税率、工资率、政府支出等数据。

综列数据有别于混合横截面数据的关键特征是，同一横截面数据的数据单位（上述例子中的个人、企业或县）都被跟踪了一段特定的时期。表1.4中的数据不能被看成综列数据集的原因是，在1993年和1995年出售的房子很可能不同；如果有一些房子相同，那相同房子的数目可能太小，以至无足轻重。对比之下，表1.5则是美国150个城市有关犯罪和相关统计量的一个两年综列数据集。

表1.5有几个有趣的特征。首先，每个城市都有一个编号，从1到150。哪个城市被称为城市1、城市2等并不重要。和纯横截面数据一样，对综列数据集中横截面数据的排序无关紧要。虽然我们可以用一个数字来代替城市名，但二者都保留通常很有好处。

12 表 1.5 城市犯罪统计量的一个两年综列数据集

obsno	*city*	*year*	*murders*	*population*	*unem*	*police*
1	1	1986	5	350 000	8.7	440
2	1	1990	8	359 200	7.2	471
3	2	1986	2	64 300	5.4	75
4	2	1990	1	65 100	5.5	75
⋮	⋮	⋮	⋮	⋮	⋮	⋮
297	149	1986	10	260 700	9.6	286
298	149	1990	6	245 000	9.8	334
299	150	1986	25	543 000	4.3	520
300	150	1990	32	546 200	5.2	493

第二个有用的方面是，城市 1 的两年的数据占据了观测中的前两行。观测 3 和 4 对应于城市 2，等等。由于 150 个城市中的每一个都有两行数据，所以任何一个计量经济软件包都会把它看成是 300 个观测。这个数据集还可以被看成是将 1986 年和 1990 年的两个横截面数据混合起来后，两个年份都出现的城市所组成的。但如在第 13 章和第 14 章中将看到的那样，我们还可以利用综列数据结构，来回答那些仅把它看成混合横截面数据所不能回答的问题。

在表 1.5 中组织观测值时，我们将每个城市两年的数据依次相连，第一年的数据总是在第二年的数据之前。从几乎每个实用的角度来看，这是组织综列数据集的最好方法，可以将这种组织数据的方法与表 1.4 中存储混合横截面数据的方法进行比较。简言之，像表 1.5 那样组织综列数据的原因在于，我们需要对每个城市两年的数据进行变换。

由于综列数据要求同一单位不同时期的重复观测，所以要得到综列数据(特别是那些个人、家庭和企业的数据)，比得到混合综列数据更困难。对同样的观测单位观测一段时间应该比横截面数据甚至混合横截面数据要有一些优越性。本书中所强调的好处是，对同一单位的多次观测，使我们能控制个
13 人、企业等观测单位本身具有而我们又观测不到的特征。如将看到的那样，使用不止一次的观测，使我们在单纯横截面数据下很难作出因果推断的情况下进行因果推断。综列数据的第二个优点是，它通常使我们能研究决策行为和结果之滞后的重要性。由于预期许多经济政策在一段时间之后才产生影响，所以综列数据所反映的信息就更有意义。

多数本科水平的教材都不讨论综列数据的计量方法。但经济学家现在都认识到，没有综列数据，即便不是不可能，也很难对某些问题作出令人满意

的回答。你将会看到，借助综列数据分析这种不比处理标准横截面数据集困难多少的方法，我们能取得相当可观的进展。

对数据结构的评论

本书的第一部分首先考虑了对横截面数据的分析，因为这方面存在的概念和技术困难最少。同时，它也揭示了计量分析的多数重要主题。在本书的剩余部分，我们还会用到横截面分析所提供的方法和深刻见解。

虽然对时间序列的计量分析也用到许多与横截面分析相同的工具，但由于许多经济上的时间序列的趋势性和高度持续性，所以对它的分析更复杂。过去用来说明计量方法用于时间序列数据之方式的例子，现在都普遍被认为是无效的。由于首先就使用这种例子只会使落后的计量实践进一步加深，所以这种做法没有意义。于是，我们把对时间序列的处理推后到第二部分，到那时还将引入有关趋势、持续性、动态性和季节性等问题。

在第三部分，将详尽分析混合横截面数据和综列数据。对独立的混合横截面数据和简单的综列数据的分析，基本上就是对纯横截面数据分析的直接推广。尽管如此，我们还是等到第13章才讨论这些专题。

1.4 计量经济分析中的因果关系和其他条件不变的概念

在多数对经济理论的检验中（当然包括对公共政策的评价），经济学家的目标就是要推定一个变量对另一个变量（如犯罪率或工人的生产力）具有**因果性效应**（causal effect）。虽然简单地发现两个或多个变量之间有某种联系很诱人，但除非能得到某种因果关系，否则这种联系很难令人信服。

其他条件不变（ceteris paribus）——意味着“其他（相关）因素保持不变”——的概念在因果分析中有重要作用。虽然这种思想早已暗含在我们前面的某些讨论之中，特别是例1.1和例1.2，但迄今为止我们还没有明确提出。

你可能记得，在初级经济学教程中，很多经济学问题都有其他条件不变的特征。比如，在分析消费需求时，我们想知道一种商品价格的变化对其需
14 求量的影响，而让所有其他因素——譬如收入、其他商品的价格和个人嗜好等——都保持不变。如果不保持其他因素不变，我们就不可能知道价格变化对需求量的因果性效应。

保持其他因素不变的做法对政策分析也至关重要。在在职培训的例子（例1.2）中，我们可能对在其他成分（特别是受教育程度和工作经历）不变的情况下多一周在职培训对工资的影响感兴趣。如果我们能保持所有其他相关因素不变，然后发现在职培训与工资之间有一种联系，那么就能断定在

职培训对工人的生产力具有因果性效应。尽管看上去相当简单，但即便在现阶段也应该清楚，除非在极为特殊的情形下，否则不可能如实地保持所有其他因素不变。多数经验研究中的一个关键问题是：对于作出因果推断而言，是否有足够多的其他因素被保持不变？对一项计量经济研究进行评价时，都要面临这个问题。

在多数重大的应用中，能影响我们所关注的变量——如犯罪活动或工资——的因素为数众多，而想要隔离任何一个特定的变量看来都像是无谓的努力。但最终我们将看到，只要善加利用，用计量经济方法就可以模拟一个其他条件不变的实验。

就此看来，我们还不能解释如何使用计量经济方法去估计那种保持其他条件不变的影响，所以我们将考虑在作出经济学中的因果性推断时可能出现的某些问题。在这种讨论中，我们不会使用任何方程。对于每一个例子，如果能进行适当的实验，推断因果关系的问题便不复存在。因此，不妨设想一下怎样才能构造一个这样的实验，并从而观察到，在多数情况下，要得到实验数据的想法是不切实际的。但考虑一下，为什么可供使用的数据不具备一个实验数据集的重要特征，也是有好处的。

迄今为止我们都只是靠直觉来理解诸如随机、独立和相关等术语，而这些在初级的概率论和数理统计教程中都有详细介绍。（附录 B 回顾了这些概念。）我们先举一个例子，来说明这些重要问题。

例 1.3　肥料对作物收成的影响

一些早期的计量研究［例如格里利奇（Griliches，1957）］考虑了新的肥料对谷物收成的影响。假设所考虑的作物是大豆。由于施肥量只是影响收成的因素之一——其他因素还包括降雨量、土壤质量和攀附植物的出现等，所以这个问题必须在其他条件不变的情况下进行研究。决定施肥量对大豆收成之因果性效应的途径之一，是按如下步骤进行一项实验。选择几块一亩大的土地，再对每块土地施加不同数量的肥料，然后度量每块土地的收成，这就形成一个横截面数据。于是，使用统计方法（在第 2 章将介绍）就可以测度收成与施肥量之间的关系。

如前所述，这个实验看起来并不是很好，因为我们并没有提到在选择实
15 验田时要保证除施肥量不同外其他方面都要完全一样。实际上，选择这种试验田是不可行的：诸如土壤质量等某些其他因素甚至都不能被充分观测到。我们如何知道这个实验的结果可用来度量施肥量在其他条件不变的情况下的影响呢？答案取决于如何选择施肥量的详细情况。如果在确定每一块田的施肥量时施肥量的选择独立于影响收成的其他土地特征——即在决定施肥量时（可以）* 完全忽略土地的其他特征，那么我们就可以达到目标。在第 2 章，我们将证明这种观点的正确性。

---

* 原文有语病，“可以”二字是译者添加的。——译者注

对于应用经济学中推断因果关系时出现的困难，下面这个例子更具代表性。

例 1.4　测度教育的回报

劳动经济学家和政策制定者长期以来都对“教育的回报”有浓厚兴趣。这个问题可以多少有些不太规范地这样提出：如果从总体中选择一个人，并让他或她多受一年教育，那么，他或她的工资会提高多少？像前面的例子一样，这也是一个其他条件不变的问题，即意味着在保持所有其他条件不变的情况下，对这个人增加一年的教育。

我们可以设想有一位社会计划者在设计一项实验以回答这个问题，这与农业研究者设计一项实验来估计施肥量的影响很相像。模仿例 1.3 中施肥量实验的一种方法是：选择一群人，随机地给每个人一定的教育程度（给某些人 8 年级的教育，给某些人高中教育等），然后度量他们的工资（假设这时每人都有工作）。这里的人就像上例中的实验田，教育则起到肥料的作用，而工资则类似于大豆的收成。像例 1.3 一样，如果受教育水平的确定独立于影响生产力的其他特征（如工作经历和天生能力），那么忽略这些其他因素的分析将会得到有用的结论。同样，第 2 章将花点工夫说明这个观点。到目前为止，我们只是在没有任何根据的情况下这么说。

与肥料—收成的例子不同，例 1.4 中描述的实验是不可行的。不用说经济上的成本，要对一群人随机地决定其受教育水平，在道义上也明显成问题。作为一个逻辑上的问题，如果一个人已有大学文凭，我们就不可能只给他 8 年级的教育。

尽管不能得到度量教育回报的实验数据，但我们肯定可以从工作人群总体中随机地抽取一大群人，并搜集其教育水平和工资的非实验数据，但这些
16 数据所具有的特点，使它很难用于估计教育在其他条件不变情况下的回报。人们选择其自身受教育的水平，因此受教育水平可能并不独立于所有那些影响工资的因素。这是大多数非实验数据集所共有的一个特点。

影响工资的一个因素是其工作经历。由于追求更多的教育通常需要推迟其加人劳动力队伍的时间，所以那些受教育越多的人工作经历就越短。因此，在一个有关工资和受教育水平的非实验数据集中，受教育水平可能与影响工资的一个关键变量呈负相关。还要知道，一个人越有天分，就越会选择接受更多的教育。由于越大的能力导致越多的工资，所以我们又发现受教育水平与影响工资的一个重要因素呈相关关系。

在例 1.4 中省略的因素，工作经历和个人能力，在例 1.3 中也有相似的因素。工作经历通常易于度量，因此与降雨量相似。另一方面，能力则含糊不清而又难以量化，所以它与例 1.3 中的土壤质量相似。像在我们本书中将要看到的那样，在估计教育这种变量在其他因素不变情况下的影响时，分离出像工作经历这种可观测度量的影响相对简单。我们还将发现，要分离像个人能力这种本来就难以观测的因素的影响则相当成问题。公平地讲，计量经

济方法的许多进展，都试图对计量模型中的这些无法观测因素进行处理。

例1.3和例1.4之间可以得到一个可比拟之处。试想，在例1.3中，施肥量并非完全随机决定。相反，选择施肥数量的助手认为，最好在较高质量的田里施更多的肥料。（尽管这些农业研究者不能对这些土地之间的差别充分量化，但他们应该大致知道哪一块土地的质量较好。）这种情形，和例1.4中受教育水平与无法观测的个人能力相关的情况完全类似。因为较好的土地会导致较好的收成，而在较好的土地上又使用了较多的肥料，所以观察到的收成与用肥量之间的关系可能是有欺骗性的。

例1.5　执法对城市犯罪活动的影响

如何最好地制止犯罪的问题已伴随了我们一段时间，并将继续成为我们考虑的问题。这方面一个特别重要的问题是：更多的警官出现在大街上会制止犯罪吗？

这个其他条件不变情况下的问题很容易表述：如果随机地选择一个城市并多增加10名警官，那么犯罪率会下降多少？另一种表述这个问题的方法是：如果两个城市各方面都相同，只是A市比B市多10名警官，那么这两个城市的犯罪率会有多大的差别呢？

要找到除警力规模外完全一样的两个社区根本不可能。幸运的是，计量经济分析并不要求这样。我们真正需要知道的是，我们所搜集到的有关社区犯罪水平和警力规模的数据是否可被视为实验数据。我们确实可以设想一个真正的实验，其中有许多城市，而每个城市在下一年度将使用多少警力则完全由我们控制。

尽管可以利用政策来影响警力规模，但我们显然不能告诉每个城市该雇
17 用多少警官。如果（很可能）一个城市有关雇用多少警官的决策与另一个城市影响犯罪的因素相关，那么，这些数据就必须被看做是非实验性数据。实际上，看待这个问题的方法之一是，将一个城市对警力规模的选择与该城市的犯罪量看成**同时决定的**（simultaneously determined）变量。我们将在第16章清楚地解决这种问题。

我们前面讨论过的三个例子已探讨了各个层面（如个人或城市等层面）的横截面数据。在时间序列问题中进行因果推断时也会出现同样的障碍。

例1.6　最低工资条例对失业的影响

一个重要但可能有争议的政策问题是，最低工资条例对各种工人群体的失业率有什么影响？尽管这个问题可以在各种数据背景（横截面数据、时间序列数据或综列数据）下进行研究，但通常都是使用时间序列来考察其总影响的。表1.3给出了失业率和最低工资方面时间序列数据集的一个例子。

标准的供求理论意味着，当最低工资被提高到市场出清工资之上时，劳动需求便会沿着劳动需求曲线向上移动，从而导致总就业减少。（劳动供给大于劳动需求。）为了定量分析这种影响，我们可以研究就业与最低工资之

间在不同时间的关系。除了处理时间序列数据可能引起某些特殊问题之外，在进行因果推断时可能还有一些问题。美国的最低工资并非在隔绝状态下确定，所以各种经济和政治力量对任一给定年份最终的最低工资都有影响。（一旦确定了一个最低工资，通常在几年内都不会改变，除非相对通货膨胀而指数化。）因此，最低工资的大小很可能与影响就业水平的其他因素有某种联系。

我们可以设想，美国政府在实施一项实验，以决定最低工资对就业的影响（以消除人们对低工资工人福利的担忧）。政府每年可以随机地设定最低工资，然后就业结果可以被制成表格。于是，使用相当简单的计量分析方法，就可以对如此得到的实验性时间序列数据进行分析。但这种方案无法描述最低工资是如何设定的。

如果我们能充分控制其他与就业相关的因素，那我们就仍有希望去估计最低工资在其他条件不变情况下对就业的影响。在这个意义上，这个问题与前面横截面数据的例子极为相似。

即便经济理论不能很自然地用因果关系进行描述，使用计量经济方法通常也能对它们的预测进行检验。如下即是这种方法之一例。

例 1.7　预期假说

18 来自金融经济学的预期假说认为，给定投资者在进行投资决策时可供使用的全部信息，任何两种投资的期望回报都相同。比如，考虑两个可能的投资，投资的时间长度为 3 个月，并在同时进行：(1) 购买一份面值 10 000 美元 3 个月期国库券，其价格低于 10 000 美元；到了 3 个月，你就得到这 10 000 美元。(2) 购买一份 6 个月期国库券（价格低于 10 000 美元），到了 3 个月，把它作为一份 3 个月期国库券售出。虽然每种投资所需要的初始资本大致相当，但二者之间存在重大差别。对于第一项投资，由于你不仅知道 3 个月期国库券的面值，而且还知道其价格，所以你在购买时就确切地知道其回报。对第二项投资则不然：虽然你在购买时就知道 6 个月期国库券的价格，但你并不知道在 3 个月后能卖个什么样的价格。因此，对于一个投资期限只有 3 个月的人来说，这种投资具有不确定性。

这两种投资的实际回报通常存在差异。根据预期假说，给定投资时的全部信息，第二项投资的预期收益应该与购买 3 个月期国库券的收益相同。像我们将在第 11 章看到的那样，这个理论相当容易检验。

▶ 小　结

在这个绪论中，我们已讨论了计量经济分析的目的和研究范围。所有的

应用经济学领域都使用计量经济学来检验经济理论、指导政府和私人政策制定者，以及预测经济时间序列。虽然计量经济模型有时也可以从规范的经济模型推导出来，但在另一些情况下，计量经济模型都是以非规范的经济推理和直觉为基础的。任何一个计量经济分析的目的都是估计模型中的参数并检验对这些参数的假设；参数的数值和符号，决定了经济理论的有效性和某项政策的效果。

横截面数据、时间序列数据、混合横截面数据和综列数据，是应用计量经济学中最经常使用的数据结构类型。由于多数经济时间序列在时间上的相关关系，所以对牵涉到时间维的数据集（如时间序列和综列数据），都需要特别处理。另外，诸如趋势性和季节性等问题，也只会在对时间序列数据的分析中出现，横截面数据则不存在这种问题。

在1.4节，我们讨论了其他条件不变和因果性推断的概念。多数情况下，社会科学中的假设都具有“其他条件不变”的特点：在研究两个变量之间的关系时，所有其他的相关因素都必须固定不变。因为社会科学中所搜集到的多数数据都具有非实验特征，所以发现其中的因果关系极具挑战性。

关键术语

19 因果性影响
其他条件不变
横截面数据集
数据频率
计量经济模型
经济模型
经验分析
实验数据
非实验数据
观测数据
综列数据
混合横截面数据
随机抽样
时间序列数据

经济科学译丛·计量经济学导论·现代观点　经济科学译丛·计量经济学导论·现代观点　经济科学译丛·计量

第1篇

横截面数据的回归分析

21 在本书的第 1 篇，我们要讨论横截面数据的回归分析。高等代数和概率论与数理统计的基本概念的坚实基础，对本篇知识的学习是必不可少的，所以我们会在附录 A，B 和 C 中完整地复习这些内容。在第 2 章中，我们从用一个变量去解释另一个变量的简单线性回归模型开始。虽然简单回归在应用计量经济学中并非广泛使用，但因其涉及的代数知识及其释义都相对简单明了，所以有时也会被使用，并自然而然地成为计量经济学的起点。

在第 3 章和第 4 章，我们将介绍多元回归分析，在多元回归分析中，容许一个以上的变量影响我们所要解释的变量。在经验研究中，多元回归仍然是应用得最为广泛的方法，所以这两章的内容值得我们仔细研究。在第 3 章，我们把重点放在普通最小二乘（OLS）的数学工具上，并建立 OLS 估计量无偏和最优线性无偏的条件。第 4 章则主要讨论统计推断中的重要论题。

在第 5 章讨论了 OLS 估计量的特性，如大样本或渐近性质。如果回归模型的误差是非正态分布的，那么这些性质就为第 4 章的推断过程提供了合理性。在第 6 章研究了回归分析的其他内容，包括高级的函数形式问题、数据度量、预测和拟合优度等。在第 7 章则解释了如何在多元回归模型中包含定性信息的问题。

在第 8 章说明了如何检验并纠正误差项中的异方差或非恒常方差的问题。我们会介绍如何校正通常的 OLS 统计量，并给出 OLS 的一种推广，即平时所说的**加权最小二乘法**，从而得以明显地考虑误差项中的不同方差。我们在第 9 章将更深入地研究一个重要问题，那就是误差项与一个或多个解释变量之间的相关性。我们将说明用代理变量解决被忽略变量问题的有效性。另外，还将证明 OLS 估计量在变量存在测量误差时的偏误和不一致性。我们还要讨论各种不同的数据问题，包括异常值的问题。

第 2 章　简单回归模型

22　简单回归模型可以用来研究两个变量之间的关系。我们会发现，出于某些原因，简单回归模型要作为经验性分析的一般工具还存在着局限性，但是在某些情况下把它当做经验工具来使用还是非常适宜的。学会解释简单回归模型，对于我们接下来几章要学习的多元回归模型，无疑也是非常好的练习。

2.1　简单回归模型的定义

许多应用计量经济学分析都是从如下假设前提开始的：y 和 x 是代表某一个总体的两个变量，我们感兴趣的是用 x 来解释 y，或者说是研究 y 如何随 x 而变化。我们在第 1 章中讨论了一些例子，其中包括：y 是大豆的产出，x 是化肥的用量；y 是每小时的工资，x 是受教育的年数；y 是社区的犯罪率，x 是警察的数量。

在写出用 x 解释 y 的模型时，我们要面临三个问题。首先，既然两个变量之间没有一个确切的关系，那么我们应该如何考虑其他影响 y 的因素呢？第二，y 和 x 的函数关系是怎样的呢？第三，我们怎样知道我们是否抓住了

在其他条件不变的情况下 y 和 x 之间的关系（如果这是我们所追求的目标的话）呢？

我们可以通过写出关于 y 和 x 的一个方程来消除这些疑惑。一个简单的方程是

$$y=\beta_0+\beta_1 x+u \tag{2.1}$$

且假定方程（2.1）在我们所关注的某个总体中成立，它定义了一个**简单线性回归模型**（simple linear regression model）。因为它把两个变量 x 和 y 联系起来，所以又把它叫做**两变量**或者**双变量线性回归模型**。我们现在来讨论等式（2.1）中每个量的含义。［“回归”这个词的起源对于大部分现代计量经济学的应用来说并不特别重要，所以我们在此就不予讨论。参见斯蒂格尔（Stigler，1986）对回归分析充满魅力的历史所做的介绍。］

23 如果和方程（2.1）联系起来，变量 y 和 x 就有许多可以互换的不同名称。y 可叫做**因变量**（dependent variable）、**被解释变量**（explained variable）、**响应变量**（response variable）、**被预测变量**（predicted variable）或者**回归子**（regressand）。x 则被叫做**自变量**（independent variable）、**解释变量**（explanatory variable）、**控制变量**（control variable）、**预测元**（predictor variable）或者**回归元**（regressor）。［x 还被叫做**协变量**（covariate）。］“因变量”和“自变量”两词在计量经济学中使用较多，但要注意这里所说的“自变”（independent）与统计学概念里面随机变量之间的独立（independence）有所不同（见附录 B）。

“被解释”和“解释”变量这两个词可能是最具描述性的。“响应”和“控制”在实验性科学中运用最多，因为 x 变量是在实验者的控制之下的。我们不使用“被预测变量”和“预测元”这样的字眼，尽管有时候你还是会看到它们。在表 2.1 中，我们总结了简单回归中的术语。

表 2.1　　　　简单回归的术语

y	x
因变量	自变量
被解释变量	解释变量
响因变量	控制变量
被预测变量	预测变量
回归子	回归元

变量 u 被叫做关系式中的**误差项**或者**扰动项**，表示除 x 之外影响 y 的因素。一个简单回归分析能够有效地处理除 x 之外其他所有影响 y 的非观测因素。你也可以把 u 看做“观测不到的”因素。

等式（2.1）同样表述了 y 和 x 之间的函数关系。如果 u 中的其他因素被看做是保持不变的，就意味着 u 的变化为零，即 $\Delta u=0$，那么 x 对 y 具

有线性影响，其表述如下：

$$\text{如果}\quad \Delta u = 0,\ \text{那么}\quad \Delta y = \beta_1 \Delta x \tag{2.2}$$

因此，y 的变化量是 β_1 和 x 的变化量的简单乘积。这就是说，保持 u 中其他因素不变，β_1 就是 y 和 x 的关系式中的**斜率参数**（slope parameter），在应用经济学中，它是人们研究的主要兴趣所在。**截距参数**（intercept parameter）β_0 也有它的作用，但很少被当做分析研究的主要部分。

例 2.1 大豆产出和施肥量

24 假使大豆的产出由以下模型所决定：

$$yield = \beta_0 + \beta_1 fertilizer + u \tag{2.3}$$

y = 产出，而 x = 施肥量。农业研究者对其他因素不变时化肥用量如何影响大豆产出量感兴趣。影响的效果由 β_1 给出，误差项 u 包括了诸如土壤质量、降雨量等因素。系数 β_1 度量了在其他条件不变的情况下施肥量对产出量的影响：$\Delta yield = \beta_1 \Delta fertilizer$。

例 2.2 一个简单的工资方程

以下模型表示一个人的工资水平与他的可测教育水平及其他非观测因素的关系：

$$wage = \beta_0 + \beta_1 educ + u \tag{2.4}$$

如果工资和教育分别以每小时美元数和受教育的年数来计量，那么 β_1 度量了在其他条件不变的情况下每增加一年教育所获得的小时工资增长量。其他非观测因素则包括劳动力的经验、天生的素质、在现任雇主之下供职的时间、工作道德以及无数的其他因素。

等式（2.1）的线性性显示，不管 x 的初始值为多少，它的任何一单位变化对 y 的影响都是**相同的**。这对许多经济学应用来说是非常不现实的。举例来说，在工资—教育的例子中，我们或许还要考虑到**递增的**回报，就是说，后一年的教育比前一年的教育对工资的影响**更大**。我们将在 2.4 节中研究如何考虑这种可能性。

最困难的问题是，模型（2.1）是否真的能让我们得到关于 x 如何在其他条件不变下影响 y 的结论。从等式（2.2）我们可以看到，保持 u 中的其他所有条件不变，β_1 确实能够度量 x 对 y 的影响。但我们对这个因果问题的讨论可以就此结束吗？非常不幸，还不行。一般地说，我们怎样能在保持其他因素固定的同时又忽略所有这些其他因素，以得到 x 对 y 在其他条件不变下的影响呢？

正如我们将要在 2.5 节中看到的，只有当我们对非观测的 u 与解释变量 x 之间的关系加以约束时，才能从一些数据的随机样本中获得 β_0 和 β_1 的可靠估计量。没有这样一个约束，我们就不能估计出在其他条件不变下的影

响 β_1。因为 u 和 x 都是随机变量，所以我们需要一个基于概率的概念。

在我们介绍关于 x 和 u 的关系的这个关键的假定之前，我们可以先作出一个关于 u 的假定。只要截距 β_0 被包括在等式之中，假设总体中 u 的平均值为零就不会失掉什么。

25 用数学形式来表示就是

$$\mathrm{E}(u)=0 \tag{2.5}$$

非常重要的是，假定（2.5）并没有说出 u 和 x 的关系，只是简单地说明了总体中非观测变量的分布。用前面的例子来说，我们可以看到，假定（2.5）的约制性并不是特别强。在例 2.1 中，我们把诸如土壤质量这样的对大豆产出有影响而观测不到的因素进行标准化，使其在所有耕种的地区平均值为零，对结果不会有损失。例 2.2 中非观测因素的情形也与此相同。为不失一般性，我们可以假定在所有的工作人员构成的样本中诸如平均能力这样的因素为零。如果你还不信服，试做一下习题 2.2，你就会发现，我们总能够通过重新定义等式（2.1）中的截距使得假定（2.5）成立。

我们现在来看看关于 u 和 x 的关系的关键性假定。测度两个随机变量的关系的非常自然的方法是**相关系数**（定义和性质见附录 B）。如果 u 和 x **不相关**，那么作为随机变量，它们就没有**线性**关系。为了界定方程（2.1）中的 u 和 x 没有关系而作出 u 和 x 不相关（或没有相关关系）的假定，虽然迈出了一大步，但还走得不够远。这是因为相关关系只是度量 u 和 x 之间的线性相依性。而相关关系有着与人们的直觉相违的性质，如 u 与 x 不相关，但是却可能与 x 的函数比如说 x^2 相关。（参见 B.4 节的进一步讨论。）对于大部分做回归的目的来说，这种可能性是不可接受的，因为它会在解释模型和推导统计学性质时出现问题。一种更好的方法是对**给定 x 时 u 的期望值**作出假定。

因为 u 和 x 是随机变量，所以我们能够在任何给定的 x 值下得到 u 的条件分布。具体地说，对于任何一个 x 值，我们都能够在 x 的值所描述的总体剖面上求得 u 的期望（或平均）值。关键的假定是，u 的平均值不依赖于 x 值。我们可以把它写作：

$$\mathrm{E}(u|x)=\mathrm{E}(u)=0 \tag{2.6}$$

其中，第二个相等关系来自于式（2.5）。等式（2.6）中的第一个相等关系是一个新的假定，叫做**零条件均值假定**（zero conditional mean assumption）。这就是说，对任何给定的 x 值，非观测因素的均值是相等的，因此它们必须与整个总体中的 u 的均值相等。

让我们来看一下在工资的例子中（2.6）意味着什么。为使讨论简化，令 u 为天生能力。那么，式（2.6）就要求不管受教育的年数为多少，平均能力水平都是一样的。例如，如果 $\mathrm{E}(abil|8)$ 表示所有受过 8 年教育的人的

26 平均能力，$\mathrm{E}(abil|16)$ 表示所有受过 16 年教育的人的平均能力，那么式（2.6）就意味着这两者是相同的。事实上，对所有教育水平的人来说，平均能力都必定是相等的。如果，比方说我们认为平均能力是随着受教育年数的

增加而增长的，那么式（2.6）就是错的。（平均来说，如果越有能力的人选择接受越多的教育，那么这种情形就很有可能出现。）因为我们观察不到天生的能力，所以无法确知对所有的教育水平来说平均能力是否一样，但这是我们在应用简单回归分析之前必须提出的问题。

问题 2.1 ☞

假使期末考试的分数（*score*）决定于出勤率（*attend*）和影响考试成绩的其他非观测因素（如学生能力）：

$$score = \beta_0 + \beta_1 attend + u \tag{2.7}$$

那么这个模型能够满足式（2.6）吗？

在施肥的例子中，如果施肥量与该地区的其他条件没有关系，那么式（2.6）就能够成立，即土地的平均质量不会依赖于施肥量。然而，如果更多的肥料被施用在更高质量的土地上，那么 u 的期望值就会随着肥料的用量而改变，式（2.6）也就不成立了。

假定式（2.6）为 β_1 提供了另一种非常有用的解释。取以 x 为条件的式（2.1）的期望值，并利用 $E(u|x)=0$，得到

$$E(y|x) = \beta_0 + \beta_1 x \tag{2.8}$$

等式（2.8）显示**总体回归函数**（population regression function，PRF），$E(y|x)$ 是 x 的一个线性函数。线性性质意味着 x 增加一个单位，将使 y 的期望值改变 β_1 之多。如图 2.1 所示，对任何给定的 x 值，y 的分布都以 $E(y|x)$ 为中心。

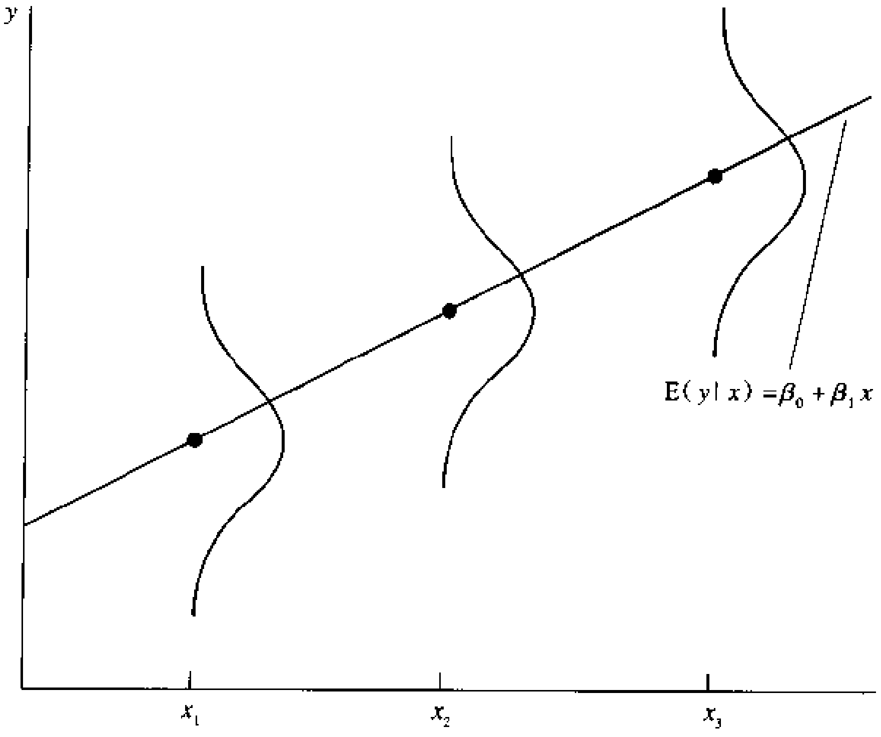


图 2.1　x 的线性函数 $E(y|x)$

27 当式（2.6）为真的时候，把 y 分成两个部分就很有用。$\beta_0 + \beta_1 x$ 有时被叫做 y 的**系统部分**——也就是说，是 y 被 x 解释的部分，和 u，即**非系统部分**，或者说是 y 不能被 x 解释的部分。我们将在下一节利用假定（2.6）促成对 β_0 和 β_1 的估计。这个假定对 2.5 节的统计学分析也非常重要。

2.2 普通最小二乘法的推导

我们已经讨论了简单回归模型的基本要素，接下来将要阐述如何估计方程（2.1）中的参数 β_0 和 β_1 这个重要问题。为了达到这个目的，我们需要从总体中找一个样本。用 $\{(x_i, y_i): i = 1, \cdots, n\}$ 表示从总体中随机取出的容量为 n 的样本。因为这些数据来自于方程（2.1），所以可以将其写为

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{2.9}$$

该式对任何 i 都成立。在这里，u_i 是第 i 次观察的误差项，它包括除了 x_i 之外的所有影响 y_i 的因素。

举例来说，在一个特定的年份里，x_i 是家庭 i 的年收入，y_i 是家庭 i 的年储蓄量。如果我们收集了 15 个家庭的数据，那么 $n = 15$。图 2.2 给出了这个数据集合的散点图和（必然为虚构的）总体回归函数。

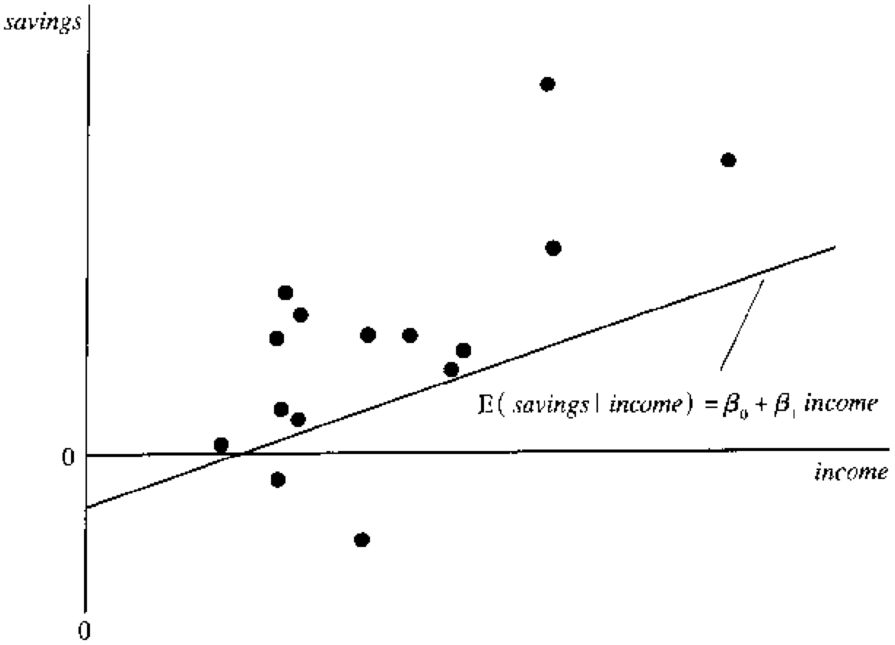


图 2.2　15 个家庭的储蓄和收入的散点图，以及总体回归
$E(savings \mid income) = \beta_0 + \beta_1 income$

在储蓄对收入的总体回归中，我们必须确定，如何才能运用这些数据来获得截距和斜率的估计值。

有几种方法可以帮助我们完成下面的估计过程。我们要用到假定（2.5）

和假定（2.6）的重要内容，即在总体中 u 为零均值而且与 x 不相关。因此，我们看到，u 的期望值为零，x 和 u 之间的协方差为零：

$$\mathrm{E}(u)=0 \tag{2.10}$$

$$\mathrm{Cov}(x,u)=\mathrm{E}(xu)=0 \tag{2.11}$$

式（2.11）中的第一个等式是随式（2.10）而来的。（协方差的定义和性质见B.4 节。）用可观测变量 x 和 y 以及未知参数 β_0 和 β_1 来表示，方程（2.10）和方程（2.11）可分别写为

$$\mathrm{E}(y-\beta_0-\beta_1 x)=0 \tag{2.12}$$

和

$$\mathrm{E}[x(y-\beta_0-\beta_1 x)]=0 \tag{2.13}$$

28 方程（2.12）和方程（2.13）指出了对总体中的（x，y）的联合概率分布的两个限制。因为要估计两个未知参数，所以我们或许会期望方程（2.12）和方程（2.13）能为我们带来关于 β_0 和 β_1 的比较好的估计量。事实上，它们确实能够做到这一点。给出一组数据，我们就能选择估计值 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 来解决方程（2.12）和方程（2.13）的样本对应问题：

$$n^{-1}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)=0 \tag{2.14}$$

$$n^{-1}\sum_{i=1}^{n}x_i(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)=0 \tag{2.15}$$

这是一个用矩法进行估计的例子。（见 C.4 节对不同的估计方法的讨论。）这两个方程可用来解出和。

利用附录 A 中求和运算的基本性质，可将方程（2.14）改写为

$$\bar{y}=\hat{\beta}_0+\hat{\beta}_1\bar{x} \tag{2.16}$$

此时 $\bar{y}=n^{-1}\sum_{i=1}^{n}y_i$ 是 y_i 的样本均值，$\bar{x}$ 类同。利用这个方程，我们可以用 $\hat{\beta}_1$，$\bar{y}$ 和 $\bar{x}$ 表示出 $\hat{\beta}_0$：

29 $$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x} \tag{2.17}$$

因此，当我们拥有了斜率估计值 $\hat{\beta}_1$ 时，给定 $\bar{y}$ 和 $\bar{x}$，自然就能够得到截距估计值 $\hat{\beta}_0$。

舍去式（2.15）中的 n^{-1}（因为它不会影响结果）并把式（2.17）代入式（2.15），得到

$$\sum_{i=1}^{n}x_i[y_i-(\bar{y}-\hat{\beta}_1\bar{x})-\hat{\beta}_1 x_i]=0$$

经重新调整，就可以得到

$$\sum_{i=1}^{n}x_i(y_i-\bar{y})=\hat{\beta}_1\sum_{i=1}^{n}x_i(x_i-\bar{x})$$

根据求和运算的基本性质［见（A.7）和（A.8）］，有

$$\sum_{i=1}^{n}x_i(x_i-\bar{x})=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

且 $$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

因此，只要有

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0 \tag{2.18}$$

估计的斜率就为

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.19}$$

方程（2.19）不外是 x 和 y 的样本协方差与 x 的样本方差的商。（见附录 C。用分子和分母同除 $n-1$，对结果不会有影响。）这是非常有意义的，因为当 $\mathrm{E}(y)=0$ 且 $\mathrm{Cov}(x, u)=0$ 时，β_1 等于总体的协方差除以 x 的方差。一个直接的暗示是，如果样本中的 x 和 y 是正相关，那么 $\hat{\beta}_1$ 为正；如果 x 和 y 是负相关，那么 $\hat{\beta}_1$ 就为负。

尽管假定（2.6）有助于得到求出式（2.17）和式（2.19）的方法，但我们在计算特定例子中的估计值时，惟一需要的假定却是式（2.18）。这几乎不能算做什么假定，因为只要样本中的 x_i 不是完全相等的，式（2.18）就一定成立。如果式（2.18）不成立，我们要么是在从总体中取样时非常不走运，要么就是没有一个值得我们关注的问题（因为 x 在总体中没有变化）。例如，令 $y = wage$，$x = educ$，如果样本中的每一个人都接受了相同年数的教育（比方说每一个人都是高中毕业生，见图 2.3），式（2.18）便不成立。只要有一个人受教育的年数不同，式（2.18）就仍然成立，并且能把 OLS 估计值计算出来。

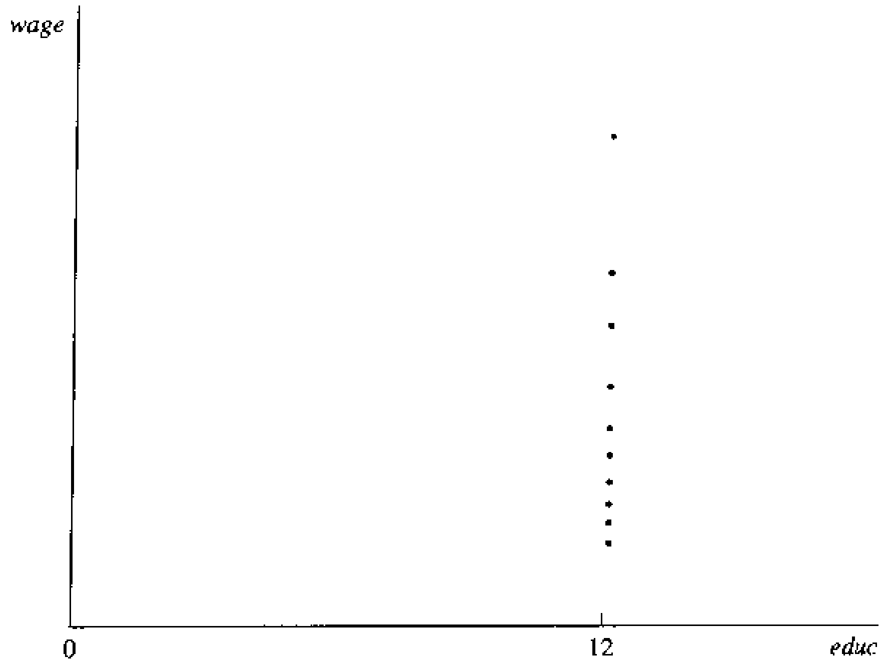


图 2.3　当对所有的 i，均有 $educ_i = 12$ 时的工资对教育的散点分布图

式（2.17）和式（2.19）所给出的估计叫做 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的**普通最小二乘**（ordinary least squares，OLS）估计。为了解释这个名称，对任何给定的截距和斜率 $\hat{\beta}_0$ 和 $\hat{\beta}_1$，当 $x = x_i$ 时，定义 y 的一个**拟合值**为

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{2.20}$$

这是当 $x = x_i$ 时我们对 y 的预测值。对样本中的每一次观测都有一个拟合值。第 i 次观测的**残差**（residual）是 y_i 的实际值和它的拟合值之差：

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \tag{2.21}$$

这样的残差有 n 个。[它们和在方程（2.9）中看到的误差不同，我们将在2.5节继续讨论这个问题。] 图2.4给出了这些拟合值和残差。

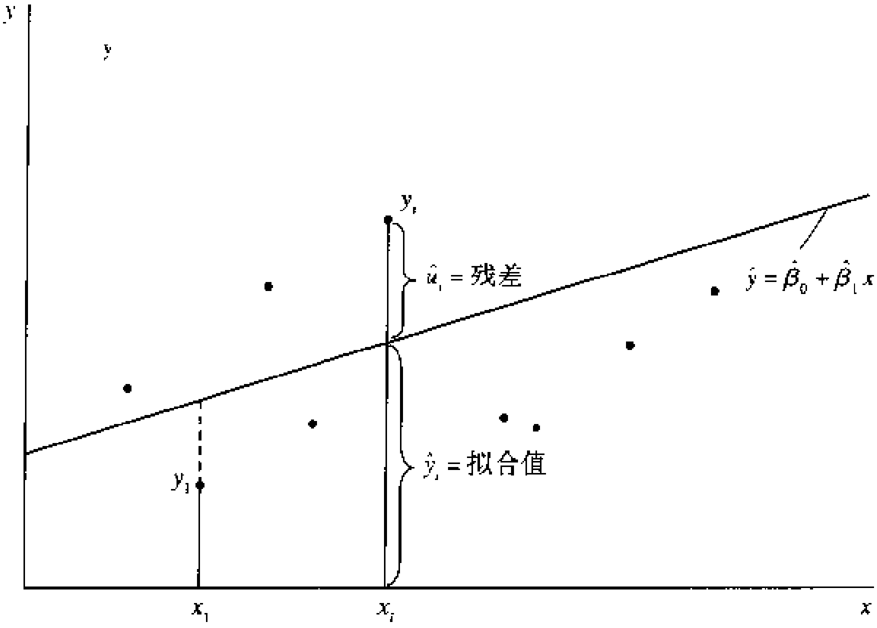


图2.4 拟合值和残差

现在，假使我们选择 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 来产生**残差的平方和**（sum of squared residuals）：

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{2.22}$$

31 并使其尽可能地小。这一章的附录告诉我们，舍去 n^{-1}，方程（2.14）和方程（2.15）就给出了使方程（2.22）最小化的（$\hat{\beta}_0$，$\hat{\beta}_1$）的必要条件。方程（2.14）和方程（2.15）通常被叫做OLS估计值的**一阶条件**（first order conditions），这一术语来自于运用微积分的最优方法（见附录A）。从前面的计算我们可以知道，OLS一阶条件的解可以由式（2.17）和式（2.19）得到。普通最小二乘法之所以得名，就是因为这些估计值使得残差的平方和最小。

一旦确定了OLS的截距和斜率的估计值，我们就建立了**OLS回归线**（OLS regression line）：

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{2.23}$$

$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 被理解为是从方程（2.17）和方程（2.19）中获得的。标记 $\hat{y}$，读作“y—帽”，强调从方程（2.23）得到的预测值是估计值。截距 $\hat{\beta}_0$ 是当 $x=0$ 时 $\hat{y}$ 的预测值，尽管在一些情况下把 x 设定为零并没有什么意义；就是说，$\hat{\beta}_0$ 本身并没有什么值得关注的地方。可是，当我们要利用式（2.23）来对不同的 x 值计算 $\hat{y}$ 的预测值时，必须考虑到计算式中的截距。方程（2.23）也被叫做**样本回归函数**（sample regression function，SRF），因为它是总体回归函数 $E(y|x)=\beta_0+\beta_1x$ 的一个样本估计。总体回归函数是固定然而未知的，切记这一点非常重要。因为样本回归函数是从一组给定的数据样本中得来的，所以新的样本会在方程（2.23）中产生不同的斜率和截距。

32 在大多数情况中，斜率估计值可写为

$$\hat{\beta}_1=\frac{\Delta\hat{y}}{\Delta x} \qquad (2.24)$$

这有重要含义。它告诉我们当 x 变化一单位时 $\hat{y}$ 的改变量。同样地

$$\Delta\hat{y}=\hat{\beta}_1\Delta x \qquad (2.25)$$

给定一个 x 的变化值（无论正负），都可以通过这个式子计算出 y 的预期变化。

我们现在给出一些简单回归的例子，这些例子都是通过实际数据得到的。换言之，我们要通过方程（2.17）和方程（2.19）来得到截距和斜率的估计值。因为这些例子包括许多组观测数据，所以我们要用计量经济学的软件来进行计算。现在，你还必须非常谨慎，不要指望能从这些回归中发现太多的东西，因为它们不一定能揭露多少因果关系。到目前为止，我们还没有涉及 OLS 的统计学性质。在 2.5 节，在我们明确对总体模型方程（2.1）施加假定之后，再来考虑其统计学性质。

例 2.3　首席执行官（CEO）的薪水和净资产回报率

对于由 CEO 构成的总体，令 y 代表年薪水（*salary*），以千美元为单位。即 $y=856.3$ 表示年薪水为856 300美元，$y=1\,452.6$表示年薪水为1 452 600美元。令 x 表示某个 CEO 的公司在过去三年里的平均净资产回报率（*roe*）。（净资产回报率被定义为纯收入占普通净资产的百分比。）例如，$roe=10$，表示净资产的平均回报率是10%。

为了研究这样度量的公司业绩和 CEO 薪水之间的关系，我们可以假定一个简单模型

$$salary=\beta_0+\beta_1 roe+u$$

斜率参数 β_1 衡量的是当净资产回报率增长 1 个百分点时以千美元计算的年薪水的变化。因为较高的 *roe* 对公司有好处，所以我们认为 $\beta_1>0$。

数据集 CEOSAL1.RAW 含有 1990 年 209 位 CEO 的信息；这些数据是从 *Business Week*（5/6/91）中获得的。在这个样本中，CEO 的平均年薪水是1 281 120美元，最低值和最高值分别为223 000美元和14 822 000美元。

1988 年、1989 年和 1990 年的平均净资产回报率是 17.18%，最低值和最高值分别为 0.5%和 56.3%。

利用 CEOSAL1.RAW 中的数据，*salary* 对 *roe* 的 OLS 回归线是

$$\widehat{salary} = 963.191 + 18.501\,roe \tag{2.26}$$

33 在这里，截距和斜率都被四舍五入到小数点后三位；用 $\widehat{salary}$ 来表示这是一个估计的方程。我们如何来解释这个方程呢？首先，如果净资产的回报率为零，即 $roe = 0$，那么薪水的预测值等于截距 963.191，因为薪水用千美元来计量，所以又等于963 191美元。其次，我们能够把薪水的预期变化看做 *roe* 的变化的函数：$\Delta\widehat{salary} = 18.501(\Delta roe)$。这意味着如果净资产的回报率增加 1 个百分点，即 $\Delta roe = 1$，那么薪水的预期变化就是 18.5，或者说是 18 500美元。因为式 (2.26) 是一个线性方程，所以所估计的变化与初始薪水无关。

利用式 (2.26)，我们可以很轻易地比较出 *roe* 取不同值时薪水的预测值。假设 $roe = 30$，那么 $\widehat{salary} = 963.191 + 18.501(30) = 1\,518.221$，超过 150 万美元。然而，这并不是说在一个 $roe = 30$ 的公司中某个特定的 CEO 可以赚1 518 221美元，因为还有其他许多因素会影响薪水。这只是我们从 OLS 回归线 (2.26) 得到的预测值。图 2.5 画出了这条估计线，连同一起的还有总体回归函数$\mathrm{E}(salary \mid roe)$。我们不可能得知 PRF 的真正形状，所以无法确知 SRF 有多接近 PRF。利用另外一组数据可以得到一条不同的回归线，这条回归线有可能更接近总体回归线，也有可能会离总体回归线更远。

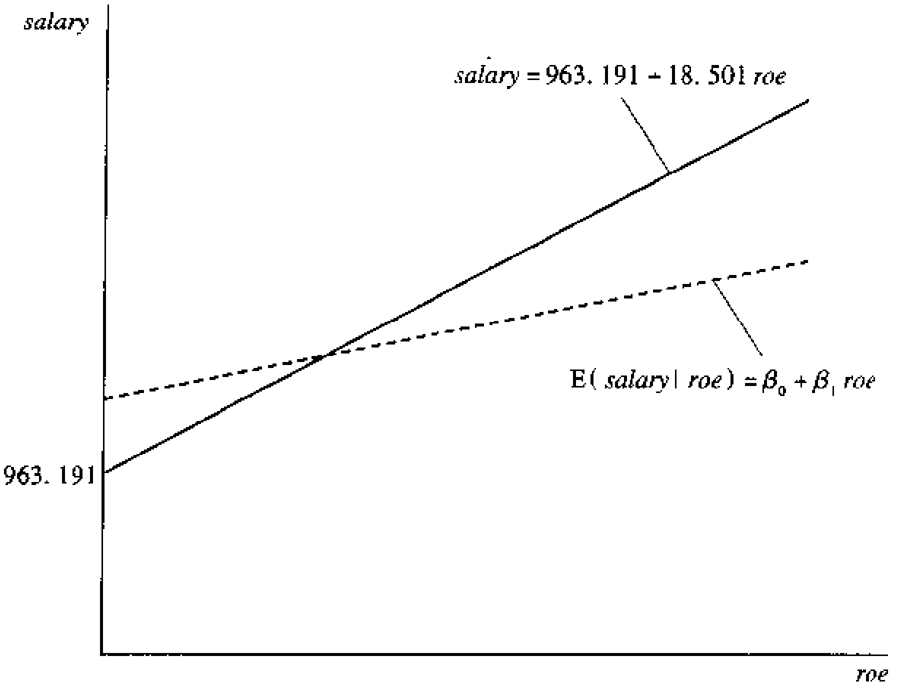


图 2.5　OLS 回归线 $\widehat{salary} = 963.191 + 18.501\,roe$ 和未知的总体回归函数

34

例 2.4　工资和教育

以 1976 年的劳动力为总体，令 y = 工资，以每小时的美元数计算。于是，对一个特定的人来说，如果 $wage = 6.75$，则表示每小时工资为 6.75 美元。令 x = 教育，表示受教育的年数。例如，$educ = 12$ 表示已完成了高中教育。由于例子中的平均工资是 5.90 美元，消费价格指数表明这一数值相当于 1997 年的 16.64 美元。

利用 WAGE1.RAW 中的数据，$n = 526$，得到下面的 OLS 回归线（或样本回归函数）：

$$\widehat{wage} = -0.90 + 0.54educ \tag{2.27}$$

我们必须谨慎地解释这个方程。截距 -0.90 从表面上看表示一个没有接受过教育的人的预测小时工资是 -90 美分。这无疑是非常可笑的。在这个样本中没有一个人接受的教育是少于 8 年的，这有助于解释零教育值的预测的不明智性。对于一个受过 8 年教育的人来说，预测工资为 $\widehat{wage} = -0.90 + 0.54(8) = 3.42$，这是用 1976 年美元表示的 3.42 美元。

式（2.27）中的斜率估计值显示，多接受一年教育，小时工资就会增加 54 美分。因此，增加四年教育可以使每小时工资增加 $4(0.54) = 2.16$ 或 2.16 美元。这是非常大的效应了。因为式（2.27）的线性性，所以无论初始教育水平怎样，每增加一年的教育对工资的增加值都是相同的。在 2.4 节，我们会讨论考虑解释变量具有非恒常的边际效应的方法。

问题 2.2☞

当 $educ = 8$，从式（2.27）中得到的估计工资是用 1976 年的美元表示的 3.42 美元，如果用 1997 年的美元表示，这个价值是多少？［提示：你可以从例 2.4 中得到足以回答这个问题的信息。］

例 2.5　选举结果和竞选支出

文件 VOTE1.RAW 包括了 1988 年美国众议院 173 次两党竞争的选举结果和竞选支出的数据。每次竞争有两名候选人，A 和 B。令 $voteA$ 为候选人 A 所得票数的百分比，$shareA$ 为候选人 A 在总竞选支出中所占的百分比。除了 $shareA$ 之外还有许多因素影响着竞选结果（包括候选人的素质，还可能包括 A 和 B 的美元开支总额）。然而，我们可以估计一个简单回归模型来看看，是否与竞争对手相比较，花费更多的钱能够得到更多的票数百分比。

利用这 173 次观测得到的估计方程为

$$\widehat{voteA} = 40.90 + 0.306shareA \tag{2.28}$$

这意味着，如果候选人 A 的开支在总花费中的比例增加 1 个百分点，候选人 A 就能够多得到几乎 1/3 个百分点的总票数。这是否存在着因果关系尚不得而知，但结果或许是我们所期待的。

在一些例子中，回归分析不是用来确定因果关系的，而只是简单地判断两个变量之间是正相关还是负相关，就像是标准相关分析一样。习题 2.12 就是这样的一个例子，它要求你用比德尔和哈默什（Biddle and Hamermesh, 1990）关于用于睡觉和工作的时间的数据来研究这两者之间的替代关系。

问题 2.3

35 在例 2.5 中，如果 $shareA=60$（即 60%），候选人 A 能得到的预测票数是多少？这个结果可信吗？

关于术语的注解

在大部分情况下，我们总是通过写出诸如式（2.26）、式（2.27）和式（2.28）的式子来表示运用了 OLS 来估计某个关系式。有时候，为了方便，仅指出运用 OLS 回归而不写出这些方程。当我们说做了

$$y \text{ 对 } x \tag{2.29}$$

的一个回归，或者只说将 y 对 x 回归，我们总是指已通过 OLS 获得了方程式（2.23）。因为总是求因变量对自变量的回归，所以 y 和 x 在式（2.29）中的位置指出了哪一个是因变量、哪一个是自变量。对于特定的应用，会交换 y 和 x 的角色。例如，为了得到式（2.26），求 *salary* 对 *roe* 的回归，或求 *voteA* 对 *shareA* 的回归来得到式（2.28）。

当使用诸如式（2.29）中的术语时，我们的用意总是要估计截距 $\hat{\beta}_0$ 和斜率 $\hat{\beta}_1$。这种情况对大部分应用都是适用的。有些时候，我们会假设截距为零（$x=0$ 暗示 $\hat{y}=0$）来估计 y 和 x 的关系；我们会在 2.6 节讨论这种情况。除非明确指出例外情况，一般都会将斜率与截距一同进行估计。

2.3 OLS 的操作技巧

在这一节中，我们将讨论拟合 OLS 回归线的一些代数性质。考虑这些性质的最佳途径，也许是意识到它们是指特定数据样本的 OLS 性质。可以把它们同 OLS 的统计学性质对立起来看，后者需要从估计量的抽样分布推演出来。我们将在 2.5 节讨论统计性质。

我们将要推导的一些代数性质看起来非常通俗。但是，对这些性质有所了解，能够帮助我们理解，在运用一定的手段处理数据时，例如当因变量和自变量的测量单位发生变化时，OLS 估计值和相关的统计值会发生什么改变。

拟合值和残差

36 假定已从给定的数据样本中得到截距和斜率的估计值 $\hat{\beta}_0$ 和 $\hat{\beta}_1$。给定 $\hat{\beta}_0$ 和 $\hat{\beta}_1$，能够获得每一次观测的拟合值 $\hat{y}_i$。[由方程（2.20）给出。] 由定义，$\hat{y}_i$ 的每一个拟合值都在 OLS 的回归线上。如方程（2.21）所指出的，与第 i 次观测相联系的 OLS 残差 $\hat{u}_i$ 是 y_i 与它的拟合值之间的差值。如果 $\hat{u}_i$ 是正的，则回归线低估了 y_i；如果 $\hat{u}_i$ 是负的，那么回归线高估了 y_i。对于第 i 次观测最理想的情况是 $\hat{u}_i=0$，但是大部分情况下并非每个残差都为零。换句话说，实际上没有一个数据点必须在 OLS 线上。

例 2.6 CEO 的薪水和净资产回报率

表 2.2 包括了 CEO 数据集合中前 15 次观测的列表，还有拟合值，即 *salaryhat*，以及残差，即 *uhat*。

表 2.2 **前 15 位 CEO 的拟合值和残差**

观测次数	*roe*	*salary*	*salaryhat*	*uhat*
1	14.1	1 095	1 224.058	−129.058 1
2	10.9	1 001	1 164.854	163.854 2
3	23.5	1 122	1 397.969	275.969 2
4	5.9	578	1 072.348	−494.348 4
5	13.8	1 368	1 218.508	149.492 3
6	20.0	1 145	1 333.215	−188.215 1
7	16.4	1 078	1 266.611	−188.610 8
8	16.3	1 094	1 264.761	170.760 6
9	10.5	1 237	1 157.454	79.546 26
10	26.3	833	1 449.773	−616.772 6
11	25.9	567	1 442.372	−875.372 1
12	26.8	933	1 459.023	−526.023 1
13	14.8	1 339	1 237.009	101.991 1
14	22.3	937	1 375.768	−438.767 8
15	56.3	2 011	2 004.808	6.191 895

37

前 4 位 CEO 的薪水比我们从 OLS 回归线（2.26）上得到的预测值要少；换言之，只给出公司的净资产回报率，这些 CEO 真正赚的比我们预测的少。而如同我们从正的"*u*—帽"所能看到的，第 5 位 CEO 赚的比我们从 OLS 回归线上预测的多。

OLS统计的代数性质

接下来要介绍一些关于OLS估计值的有用的代数性质和与它们相关的统计学知识。我们现在来介绍其中最重要的三条。

（1）OLS残差和并因而由此得到的残差的样本均值均为零。

数学表述为

$$\sum_{i=1}^{n}\hat{u}_i = 0 \tag{2.30}$$

这一性质无须证明，因为只要我们记得残差的定义是 $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$，就可以从OLS的一阶条件（2.14）推导出来。换言之，OLS估计值和是以残差的 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 为零（对任何数据集合都成立）来选择的。这并没有说任何一次观测 i 的残差是什么。

（2）回归元和OLS残差的样本协方差为零。这来自于一阶条件(2.15)，用残差表示可以把它写成如下形式

$$\sum_{i=1}^{n}x_i\hat{u}_i = 0 \tag{2.31}$$

因为OLS残差的样本均值为零，所以式（2.31）的左边与 x_i 和 $\hat{u}_i$ 之间的样本协方差成比例。

（3）点（$\bar{x}$，$\bar{y}$）总是在OLS的回归线上。换言之，如果我们利用方程（2.23），并用 $\bar{x}$ 替换 x，那么预测值就是 $\bar{y}$。这就是方程（2.16）告诉我们的。

38 **例 2.7　工资和教育**

对于WAGE1.RAW中的数据，四舍五入到两位小数，样本中的平均小时工资是5.90，平均教育是12.56。如果把 $educ = 12.56$ 代入OLS回归线(2.27)，得到 $\widehat{wage} = -0.90 + 0.54(12.56) = 5.882\,4$，若四舍五入到一位小数则为5.9。这两个值不完全一致的原因是我们把平均工资和教育以及截距和斜率的估计值都进行了四舍五入。如果最初没有四舍五入任何一个值，会得到更为接近的值，但这样做并没有什么价值。

把每个 y_i 写成它的拟合值与它的残差的和，为我们提供了解释OLS回归的又一方法。对于任何 i 值，有

$$y_i = \hat{y}_i + \hat{u}_i \tag{2.32}$$

从上面的特性（1），我们可知残差的平均值为零；等价地，拟合值的样本平均值 $\hat{y}_i$ 与 y_i 的样本均值相等，即 $\bar{\hat{y}} = \bar{y}$。进一步来说，特性(1)和(2)能被用来证明 $\hat{y}_i$ 和 $\hat{u}_i$ 之间的样本协方差为零。因此我们可以把OLS看做把 y_i 分成两个部分，即拟合值和残差，而且在样本中拟合值和残差是不相关的。

定义**总平方和**（total sum of squares，SST），**解释平方和**（explained sum of squares，SSE）和**残差平方和**（residual sum of squares，SSR）（也叫做剩余平方和）如下：

$$\mathrm{SST} \equiv \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{2.33}$$

$$\mathrm{SSE} \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{2.34}$$

$$\mathrm{SSR} \equiv \sum_{i=1}^{n}\hat{u}_i^2 \tag{2.35}$$

SST 是 y_i 的总的样本变异的度量；这就是说，它度量了 y_i 在样本中的分散程度。正如附录 C 所讨论的，如果用 $n-1$ 来除 SST，就得到了 y 的样本方差。类似地，SSE 度量了 $\hat{y}_i$ 的样本方差（这里利用了结论 $\bar{\hat{y}} = \bar{y}$），SSR 衡量的是 $\hat{u}_i$ 的样本变异。y 的总体变异能够由解释变异和未解释变异 SSR 的和来表示。因此

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR} \tag{2.36}$$

39 证明式（2.36）并不困难，但是它要求我们利用附录 A 中的求和运算的全部性质。过程

$$\begin{aligned}\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^{n}[\hat{u}_i + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^{n}\hat{u}_i^2 + 2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\ &= \mathrm{SSR} + 2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i - \bar{y}) + \mathrm{SSE}\end{aligned}$$

如果有以下等式

$$\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i - \bar{y}) = 0 \tag{2.37}$$

那么式（2.36）就成立。但前面我们已经声称，残差和拟合值之间的样本协方差为零，这个协方差又正是式（2.37）被所除的结果。由此，我们就得到了式（2.36）。

现在提出一些关于 SST，SSE 和 SSR 的注意事项是适宜的。在方程(2.33)、(2.34）和（2.35）中定义的三个量并没有统一的名字或缩写。总平方和写做 SST 或者 TSS，都不会有什么混淆之处。不幸的是，解释平方和有时候被叫做“回归平方和（regression sum of squares)”。如果给这个称谓一个自然的缩写，就很容易和残差平方和的缩写相混。所以有些回归软件倾向于把解释平方和叫做“模型平方和（model sum of squares)”。

更糟的是，残差平方和经常被称为“误差平方和（error sum of squares)”。这就尤其不幸了，因为在 2.5 节中我们会知道，误差和残差是不同的两个量。因此，我们一般都称式（2.35）为剩余平方和或者残差平方和。我们倾向于使用缩写 SSR 来表示残差平方和，因为这在计量经济学的

软件中使用较多。

拟合优度

迄今为止，我们还没有办法来衡量解释变量或者自变量 x 究竟多好地解释了因变量 y。如果能够计算出一个数值，可以用以概括 OLS 回归线对数据拟合得有多么好，对我们就非常有帮助。在接下来的讨论中，要时刻提醒自己，我们假设了斜率与截距是一起被估计的。

假定总平方和 SST 不为零——除非所有的 y_i 都相等这样一个极其罕见的情况出现，否则这个假定总是成立的——我们可以通过方程（2.36）两边同除 SST 得到 $1 = \text{SSE}/\text{SST} + \text{SSR}/\text{SST}$。回归的 ***R*-平方**（*R*-squared），有时又叫做**判定系数**（coefficient of determination），我们将其定义为

40

$$R^2 = \text{SSE}/\text{SST} = 1 - \text{SSR}/\text{SST} \tag{2.38}$$

R^2 是解释变异与总变异相比的比值，因此被解释成 y 的样本变异被 x 解释的部分。式（2.38）的第二个等式提供了计算 R^2 的另一种方法。

由式（2.36）可知，SSE 不可能比 SST 大，所以 R^2 的值总是在 0～1 之间。在解释 R^2 时，我们通常把它扩大 100 倍得到一个百分数，所以 $100R^2$ 是 y 的样本变异被 x 解释的部分的百分数。

如果数据点都落在同一直线上，OLS 就提供了一个完美的拟合。在这种情况下，$R^2 = 1$。一个接近于零的 R^2 值表明 OLS 给出了一个糟糕的拟合，因为 y_i 中的极少变异被 $\hat{y}_i$ 所捕获（后者全部落在 OLS 回归线上）。事实上，我们可以看到，R^2 就是 y_i 和 $\hat{y}_i$ 的样本相关系数的平方。这就是"*R*-平方"这个名称的由来。（习惯上，字母 R 用来表示一个总体相关系数的估计值，而且它在回归分析中被沿用至今。）

例 2.8　CEO 的薪水和净资产的回报率

在 CEO 薪水的回归中，我们可以得到

$$\widehat{salary} = 963.191 + 18.501\,roe \tag{2.39}$$

$$n = 209,\quad R^2 = 0.013\,2$$

为清晰起见，我们重述了 OLS 回归线的函数形式和观测的次数。利用该方程中所得 *R*-平方（四舍五入至四位小数），我们可以看到薪水变异有多少实际上是被净资产回报率所解释的。答案是：不多。在这 209 位 CEO 的例子中，该公司的净资产回报率仅解释了薪水变异的 1.3%。这意味着这些 CEO 的薪水变异的 98.7% 还悬而未决！之所以缺乏解释力并不奇怪，因为还有其他许多诸如公司和 CEO 个人特点等因素影响着薪水；这些因素在简单回归分析中必定被包括在误差中。

在社会科学中，回归方程中的 *R*-平方过低是很正常的，特别是对于横

截面分析来说更是如此。我们要在多元回归分析中更一般性地讨论这个问题，但我们有必要在这里强调，一个显著低的 R-平方值并不意味着 OLS 回归方程是没有用的。式（2.39）仍然可能是 *salary* 和 *roe* 之间在其他条件不变下的关系的良好估计；这是否正确并不直接依赖于 R-平方的大小。刚开始学习计量经济学的学生在进行回归方程评估时总是特别注意 R-平方的大小。但是在现在，要意识到，把 R-平方作为评价计量经济分析成功与否的主要准则会带来许多麻烦。

但有时候解释变量是能够解释因变量的样本变异中非常实在的一部分的。

例 2.9　选举结果和竞选支出

41 在竞选结果的方程（2.28）中，$R^2 = 0.505$。由此可知，在这个样本中，竞选支出的比例解释了超过 50% 的选举结果的变异。这是相当大的一个部分。

2.4　测量单位和函数形式

在应用经济学中，有两个重要的问题：（1）要理解改变因变量和/或自变量的测量单位将如何影响 OLS 估计值；（2）要了解如何把在经济学中使用的总体函数形式加入到回归分析中。附录 A 回顾了全面理解函数形式问题所用到的数学知识。

改变测量单位对 OLS 统计量的影响

在例 2.3 中，我们选择用千美元来计算年薪水，用百分数（而不是十进位数）来计算净资产回报率。为了理解方程（2.39）中的估计值，明确这个例子中的薪水和净资产回报率的测量单位非常关键。

我们还必须知道，当因变量和自变量的测量单位变化时，OLS 估计值的改变完全可以预料。在例 2.3 中，我们假设不用千美元而是用美元来计算薪水。令 *salardol* 为以美元计算的薪水（$salardol = 845\ 761$表示845 761）。当然，*salardol* 与用千美元计算的薪水有简单的关系，即 $salardol = 1\ 000\ salary$。我们其实不需要通过运用 *salardol* 对净资产回报率的回归来得到下面的估计方程：

$$\widehat{salardol} = 963\ 191 + 18\ 501\,roe \tag{2.40}$$

我们只要将式（2.39）中的截距和斜率扩大1 000倍即可得到式（2.40）中的截距和斜率。对方程（2.39）和方程（2.40）的解释也是相同的。让我们

看看方程（2.40），如果 *roe* = 0，那么 $\widehat{salardol} = 963\ 191$，所以预测薪水是 963 191美元［与我们从方程（2.39）得到的值相等］。进一步来看，如果 *roe* 增加 1，那么薪水的预测值就增加18 501美元，这与我们前面对方程（2.39）的分析得到的结论也相同。

问题 2.4 42 *假设薪水用百美元，而不是用千美元计算，令其为 salarhun，在 salarhun 对净资产回报率的回归中截距和斜率的 OLS 估计值是多少？*

概括来说，当因变量的测量单位改变时，我们很容易计算出截距和斜率估计值的变化。如果因变量乘了一个常数 c——意味着例子中的每一个值都乘了 c——那么 OLS 截距和斜率的估计值都扩大为原来的 c 倍。（这里假设自变量没有任何变化。）在 CEO 薪水的例子中，令 $c = 1\ 000$，即可把 *salary* 转换为 *salardol*。

我们同样可以用 CEO 薪水的例子来观察当改变自变量的测量单位时会发生什么情况。定义 $roedec = roe/100$ 为 *roe* 的十进位数等价量，从而 $roedec = 0.23$ 意味着 23% 的净资产回报率。为把重点放在改变自变量的测量单位上，我们回到原来对因变量即用千美元计算的薪水的讨论。当用 *salary* 对 *roedec* 进行回归时，得到

$$\widehat{salary} = 963.191 + 1\ 850.1\,roedec \qquad (2.41)$$

roedec 的系数是方程（2.39）中 *roe* 的系数的 100 倍。这是毫无疑问的。*roe* 变化 1 个百分点相当于 $\Delta roedec = 0.01$。从方程(2.41)可知，如果 $\Delta roedec = 0.01$，那么 $\Delta\widehat{salary} = 1\ 850.1(0.01) = 18.501$，通过方程（2.39）就可以得出这一结论。我们注意到，从方程（2.39）变化到方程（2.41），自变量被除以 100，所以沿用对方程的解释，OLS 斜率估计值就被放大 100 倍。概括而言，如果自变量被除以或乘以一个非零常数 c，那么 OLS 斜率系数也会分别被乘以或者除以 c。

在方程（2.41）中截距没有变，因为 $roedec = 0$ 仍然相当于零净资产回报率。一般来说，仅改变自变量的测量单位不会影响截距的估计。

在前面的章节中，我们定义 *R*-平方为衡量 OLS 回归拟合优度的标准。我们也可以研究当自变量或者因变量的测量单位发生改变时 R^2 的变化。不用任何代数运算，我们也可以知道结果：模型的拟合优度不会依赖于变量的测量单位。举例来说，薪水的变化量被净资产回报率解释的部分，不会依赖于薪水是用美元计算还是用千美元计算，或者净资产回报率是百分数还是用十进位数表示。这种直觉可以用数学方法证明，因为利用 R^2 的定义我们可知，R^2 事实上不因 y 或 x 的单位变化而改变。

在简单回归中加入非线性因素

到现在为止，我们已经重点讨论了因变量和自变量的**线性**关系。正如我

们在第 1 章中提到的，线性关系还不具备对所有经济学应用都适用的一般性。所幸我们可以通过合理定义因变量和自变量来把许多非线性因素轻易地包括到简单回归分析中。在这里我们要谈到在应用研究工作中可能出现的两种情况。

在阅读社会科学的应用性著作时，你经常会遇到一些回归方程，其中的因变量是以对数的形式出现的。这是怎么出现的呢？回忆工资—教育的例子，我们把小时工资对教育年数进行回归。得到斜率估计值 0.54［见方程 (2.27)］，这意味着每多接受一年教育，小时工资预计可以增加 54 美分。因为式 (2.27) 的线性性，所以 54 美分的增长可能是增加了第一年的教育，也可能是增加了第 20 年的教育；但这或许有些不合道理。

取而代之，我们假设增加任何一年的教育，工资增长的百分比都是相等的。43 模型 (2.27) 给出的不是稳定的百分比增长，因为百分比增长是决定于初始工资的。一个给出（近似）稳定百分比影响的模型是

$$\log(wage) = \beta_0 + \beta_1 educ + u \qquad (2.42)$$

$\log(\cdot)$表示自然对数。（见附录 A 对对数的复习。）特别地，如果 $\Delta u = 0$，那么，

$$\%\Delta wage \approx (100 \cdot \beta_1)\Delta educ \qquad (2.43)$$

注意我们如何把 β_1 乘以 100 来得到多增加一年教育时工资的百分比变化。因为工资的百分比变化对所增加的任何一年教育都是相等的，所以当教育年数增加时，工资变化也随之增加；换言之，模型 (2.42) 表明了**递增的**教育回报。通过对式模型 (2.42) 取幂，可以将其写为 $wage = \exp(\beta_0 + \beta_1 educ + u)$。$u = 0$ 时，该方程可由图 2.6 表示出来。

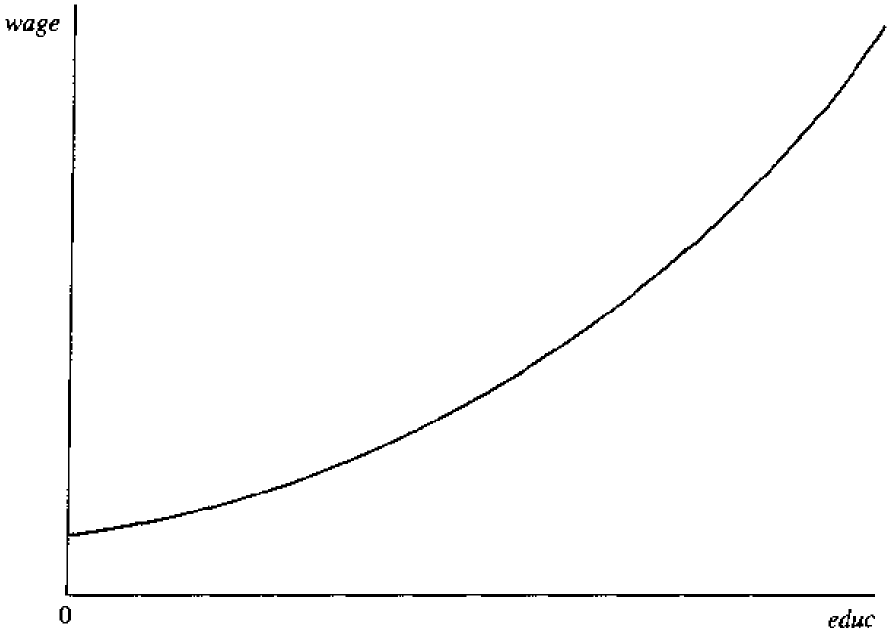


图 2.6　$wage = \exp(\beta_0 + \beta_1 educ + u)$，$\beta_1 > 0$

44 估计一个诸如式 (2.42) 的模型，用简单回归就非常直截了当。只要定义因变量 y 为$y = \log(wage)$。自变量由 $x = educ$ 来表达。OLS 的运算技巧如

前：截距和斜率的估计值由式（2.17）和式（2.19）给出。换言之，我们通过log(*wage*)对 *educ* 的 OLS 回归得到 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的值。

例 2.10　一个对数工资方程

利用与例 2.4 相同的数据，但是把log(*wage*)作为因变量，我们得到下面的关系：

$$\widehat{\log}(wage) = 0.584 + 0.083educ$$
$$n = 526, R^2 = 0.186 \tag{2.44}$$

如果将 *educ* 的系数乘以 100，它就成为一个百分数概念，表示对增加的任何一年教育，工资都会有一个 8.3% 的增长。这就是经济学家所说的“增加一年教育的回报率”。

在模型（2.42）中用对数工资主要是为了使教育对工资产生一个稳定的百分比影响，记住这一点非常重要。一旦得到方程（2.42），就几乎不用再提工资的自然对数了。特别地，如果把它说成增加一年的教育可以增加 8.3%的对数工资，就是不正确的了。

模型（2.42）中的截距没有很大的意义，它给出了当 *educ* = 0 时 log(*wage*)的预测值。*R*-平方表示教育解释了大约 18.6% 的 log(*wage*)(而不是 *wage*)的变化。最后，方程（2.44）未必不能刻画工资和所受教育之间的所有的非线性关系。例如，如果存在着“文凭效应”，那么受过 12 年的教育——即高中毕业——或许比只受过 11 年的教育有价值得多。我们要在第 7 章学习如何考虑诸如此类的非线性因素。

自然对数的另一个应用在于获得常弹性模型。

例 2.11　CEO 的薪水和公司的销售额

我们可以估计一个关于 CEO 的薪水与公司的销售额的常弹性模型。数据集与例 2.3 所使用的相同，不同的只是我们现在将薪水与销售额相联系。令 *sales* 表示公司的年销售额，并用百万美元来计算。常弹性模型如下：

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + u \tag{2.45}$$

β_1 是薪水对销售额的弹性。定义因变量为 $y = \log(salary)$，自变量为 $x = \log(sales)$，这个模型就变成了简单回归模型。用 OLS 估计这个方程得到

45

$$\log(sal\hat{a}ry) = 4.822 + 0.257\log(sales)$$
$$n = 209,\ R^2 = 0.211 \tag{2.46}$$

log(*sales*)的系数就是薪水对销售额的估计弹性。它表明公司销售额增加 1%，CEO 的薪水增加大约 0.257%——这就是我们通常对弹性的解释。

这一节中介绍的两个函数形式将要在本书后面的部分反复出现。我们讲到了包含自然对数的模型，因为它们在应用研究中出现得非常频繁。而且在

多元回归的例子中，对这些模型的解释也不会有很大的不同。

在对数形式出现的情况下，分析因变量测量单位的改变对截距和斜率估计值的影响也是非常有帮助的。因为对数形式的改变近似于一个比例的变化，所以斜率不发生变化是有道理的。对于每一次观测 i，我们可以通过把重新度量的变量记为 c_1y_i 来看到这一点。最初的方程为$\log(y_i) = \beta_0 + \beta_1x_i + u_i$。如果我们把方程两边都加一个 $\log(c_1)$，我们得到$\log(c_1) + \log(y_i) = [\log(c_1) + \beta_0] + \beta_1x_i + u_i$，或者$\log(c_1y_i) = [\log(c_1) + \beta_0] + \beta_1x_i + u_i$。（如附录 A 所述，两个对数的和等于被取对数的部分先求其乘积再取对数，这一点要记清楚。）因此，斜率仍然是 β_1，但是截距变成了$\log(c_1) + \beta_0$。类似地，如果自变量是 $\log(x)$，并且在对 x 取对数之前改变它的测量单位，则斜率保持不变，截距也不会改变。习题 2.9 会要求你证明这个结论。

我们要通过总结四个函数形式的组合来结束这一小节，既要用到它们的初始变量也要用到它们的自然对数。在表 2.3 中，x 和 y 代表变量的初始形式。以 y 为因变量，x 为自变量的模型叫做**线性到线性**模型，因为其中的每一个变量都以其线性形式出现。以$\log(y)$为因变量，x 为自变量的模型叫做**线性到对数**模型。我们在这里不会过多地讨论**对数到线性**模型，因为它在实践中很少出现。但无论如何，我们还是能在以后的章节中见到这类模型的例子。

表 2.3　　含对数的函数形式总览

模型	因变量	自变量	对 β_1 的解释
水平值—水平值	y	x	$\Delta y = \beta_1\Delta x$
对数—水平值	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
水平值—对数	$\log(y)$	x	$\%\Delta y(100\beta_1)\Delta x$
对数—对数	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1\%\Delta x$

表 2.3 的最后一列给出了对 β_1 的解释。在线性到对数模型中，$100\beta_1$ 有时被叫做 y 关于 x 的**半弹性**(semi-elasticity)。正如我们在例 2.11 中所说，在
46 对数到对数模型中，β_1 是 y 对 x 的**弹性**(elasticity)。表 2.3 值得仔细研究，我们在后面的章节中还会反复提到它。

“线性”回归的含义

在这一章中我们学习的简单回归模型又叫做简单**线性**回归模型。然而，正如我们刚刚看到的，这个一般性模型同样允许**非线性**关系的存在。那么“线

性”究竟意味着什么？你可以从方程(2.1)看到 $y=\beta_0+\beta_1x+u$。问题的关键是方程中的**参数** β_0 和 β_1 是线性的，至于 y 和 x 如何与初始的被解释变量和我们关注的解释变量相联系倒没有限制。就像我们在例 2.7 和例 2.8 中看到的，y 和 x 可以是变量的自然对数，而且这在应用中非常普遍。但我们不应该就此止步。比方说，没有什么能阻碍我们用简单回归去估计诸如 $cons=\beta_0+\beta_1\sqrt{inc}+u$ 的模型，这里 $cons$ 是年消费，inc 是年薪水。

然而简单回归模型的原理并不依赖于 y 和 x 是如何定义的，系数的解释才依赖于它们的定义。为了成功的经验性研究，善于解释系数比熟练地计算诸如式(2.19)的公式更为重要。在学习多元回归时，我们会更多地练习如何解释 OLS 回归线的估计值。

有许多模型不能被转换为线性回归模型，因为它们的参数不是线性的，例如 $cons=1/(\beta_0+\beta_1 inc)+u$。这些模型把我们带入了**非线性回归模型**的领域，但它超出了本书的范围。对大部分应用来说，选择一个能被转化到线性回归的构架之中的模型就足够了。

2.5　OLS 估计量的期望值和方差

在 2.1 节，定义总体模型 $y=\beta_0+\beta_1x+u$，并且声称，使得简单回归分析有用，一个关键假设是对于任何给定的 x 值，u 的期望值为零。在 2.2 节、2.3 节和 2.4 节中，我们讨论了 OLS 估计的代数性质。我们现在回到总体模型并研究 OLS 的**统计学**性质。换言之，把 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 看做在总体模型中出现的参数 β_0 和 β_1 的估计量。这意味着我们要研究 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 在从总体中取出的不同的随机样本中的分布性质。（附录 C 包括了一些估计量的定义和它们的一些重要性质的复习。）

OLS 的无偏性

我们从在一组简单假定的基础上建立 OLS 的无偏性开始。为方便将来参考，用简单线性回归的缩写“SLR”为字首给这些假定编号会有所帮助。第一个假定定义了总体模型。

47 **假定 SLR.1（参数的线性性）**

在总体模型中，因变量 y 与自变量 x 以及误差（干扰）u 的关系如下：

$$y=\beta_0+\beta_1x+u \tag{2.47}$$

式中，β_0 和 β_1 分别为总体的截距和斜率参数。

为使之更具真实性，y，x 和 u 都被看做在表述总体模型时的随机变量。在 2.1 节中我们仔细讨论了对这个模型的解释，并给出了一些例子。在前面

的一节里，知道方程（2.47）并不像它初看时那么严格；通过合理选择 y 和 x，能够得到有趣的非线性关系（例如常弹性模型）。

我们对用 y 和 x 的数据来估计参数 β_0，β_1，特别是 β_1 感兴趣。我们假设这些数据是作为随机样本取得的。（见附录 C 对随机抽样的复习。）

假定 SLR.2（随机抽样）

我们从总体模型中随机抽取的样本容量为 n，$\{(x_i, y_i): i = 1, 2, \cdots, n\}$。

在后面的章节讨论时间序列分析和样本选择问题时，我们将不得不指出随机抽样假定的失效。并不是所有的横截面样本都能被看做随机抽样的结果，但有许多还是可以的。

我们可以用随机样本的形式将式（2.47）写为

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \cdots, n \tag{2.48}$$

u_i 是第 i 次观测的误差项或者干扰项（例如，第 i 个人，公司 i，城市 i，等等）。因此，u_i 包括第 i 次观测中影响 y_i 的不可观测因素。不能把 u_i 和 2.3 节中定义的残差 $\hat{u}_i$ 相混淆。以后，我们会解释误差和残差之间的关系。在具体的应用中解释 β_0 和 β_1，方程（2.47）是最有启发性的，但方程（2.48）对一些统计学推理同样不可缺少。

如图 2.7 所示，关系（2.48）可用具体数据的结果绘制出来。

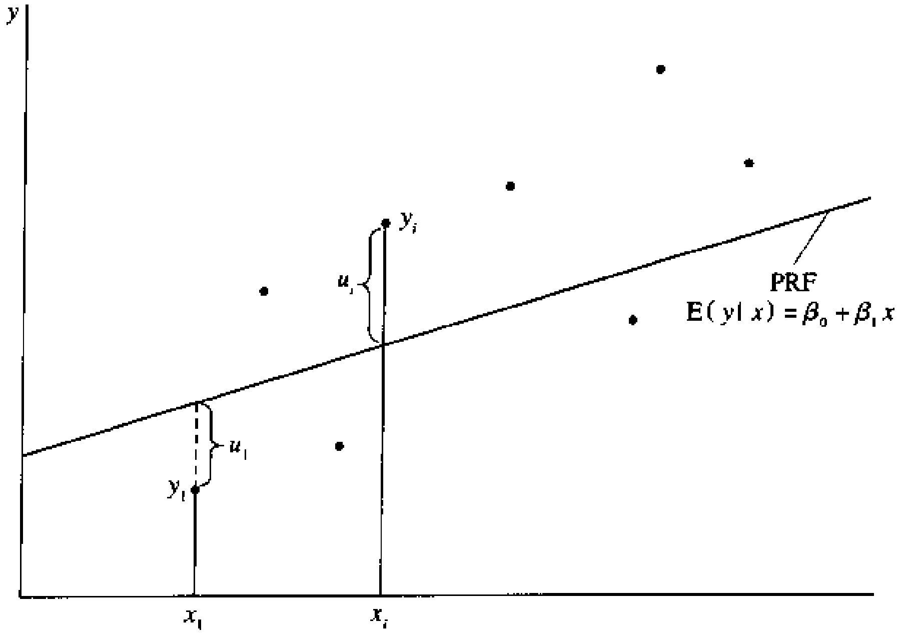


图 2.7　$y_i = \beta_0 + \beta_1 x_i + u_i$ 的图

为获得和的无偏估计量，我们必须加上 2.1 节中仔细讨论过的零条件均值假定。现在明确地把它加入我们假定的行列。

假定 SLR.3（零条件均值）

$$E(u|x) = 0$$

48 对一个随机样本来说，该假定表示 $\mathrm{E}(u_i \mid x_i)=0$，对所有的 i（$i=1, 2, \cdots, n$）成立。

除了限定总体中 u 和 x 的关系，零条件均值假定——和随机抽样假定结合起来——还为简化问题提供了方便的途径。特别地，我们可以以样本中的 x_i 的值为**条件**把 OLS 估计量的统计学性质推导出来。从技巧上来说，在统计学推理中，以自变量的样本值为条件与把 x_i 视为在**重复样本**（指屡次取样——译者注）中固定不变的处理方法相同。其过程包括以下几步：首先为 x_1，x_2，$\cdots$，x_n 选择 n 个样本值（这些值可以重复）。给定这些值，就可以得到一个关于 y 的样本（可通过获得 u_i 的随机样本而实现）。接着利用相同的 x_1，x_2，$\cdots$，x_n 以获得 y 的另一样本（样本由 n 个样本值组成——译者注）。再次利用同样的 x_i 获得 y 的又一样本，并依此重复下去。

在重复样本中不变的这种构想对非实验性的学科来说并不很现实。例如，在工资—教育一例中抽取个体时，事先选择 *educ* 的值，再在特定的教育水平下进行个体抽样，这样的想法并没有什么实际意义。（这种随机抽样代表了社会科学的经验性分析中大部分样本集合是如何获得的。）在随机抽样时，个体是随机抽取的，工资和教育都同时被记录下来。一旦假设 $\mathrm{E}(u \mid x)=0$，并且进行的是随机抽样，那么把 x_i 视为非随机在推导中就不会失掉什么。危险之处则在于，在重复样本中固定不变的假定**总是**假定了 u_i 和 x_i 是独立的。在确定何时简单回归分析将要产生无偏估计量时，用假定 SLR.3 来思考还是非常关键的。

49 当以 x_i 为条件达成一致见解时，还需要为建立无偏性的最后一个假定。

假定 SLR.4（自变量的样本有变异）

在样本中，自变量 x_i（$i=1, 2, \cdots, n$）不为相同的常数。这要求样本中的 x 要有一些变异。

我们推导 OLS 估计量的公式时会遇到假定 SLR.4，它相当于 $\sum_{i=1}^{n}(x_i-\bar{x})^2>0$。在所作出的四个假定中，这个假定的重要性是最小的，因为在人们感兴趣的应用中，它实际上都不会不成立。如果假定 SLR.4 不成立，就不能计算 OLS 的估计量，也就无从讨论它们的统计学分析。

利用结论 $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(x_i-\bar{x})y_i$（见附录 A），可以写出方程（2.19）中的斜率估计量

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2.49}$$

因为现在我们对 $\hat{\beta}_1$ 在所有可能样本中的变化感兴趣，所以 $\hat{\beta}_1$ 应被看成是随机变量。

把式（2.48）的右侧代入式（2.49），从而用总体系数和误差把 $\hat{\beta}_1$ 表示

出来。我们有

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{s_x^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta_0+\beta_1x_i+u_i)}{s_x^2}\tag{2.50}$$

为简化书写，定义 x_i 的总的变异为 $s_x^2=\sum_{i=1}^{n}(x_i-\bar{x})^2$。（这不是 x_i 的样本方差，因为没有除以 $n-1$。）利用代数求和运算，可以把 $\hat{\beta}_1$ 的分子写为

$$\begin{aligned}&\sum_{i=1}^{n}(x_i-\bar{x})\beta_0+\sum_{i=1}^{n}(x_i-\bar{x})\beta_1x_i+\sum_{i=1}^{n}(x_i-\bar{x})u_i\\&=\beta_0\sum_{i=1}^{n}(x_i-\bar{x})+\beta_1\sum_{i=1}^{n}(x_i-\bar{x})x_i+\sum_{i=1}^{n}(x_i-\bar{x})u_i\end{aligned}\tag{2.51}$$

50 如附录 A 所示，$\sum_{i=1}^{n}(x_i-\bar{x})=0$ 并且 $\sum_{i=1}^{n}(x_i-\bar{x})x_i=\sum_{i=1}^{n}(x_i-\bar{x})^2=s_x^2$。所以，可以把 $\hat{\beta}_1$ 的分子写为 $\beta_1s_x^2+\sum_{i=1}^{n}(x_i-\bar{x})u_i$。将其写在分母之上，得到

$$\hat{\beta}_1=\beta_1+\frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{s_x^2}=\beta_1+(1/s_x^2)\sum_{i=1}^{n}d_iu_i\tag{2.52}$$

这里 $d_i=x_i-\bar{x}$。现在可以看到 $\hat{\beta}_1$ 的估计量是总体的斜率 β_1 与误差 $\{u_1,u_2,\cdots,u_n\}$ 的一个线性组合相加的结果。以 x_i 的值为条件，$\hat{\beta}_1$ 的随机性完全在于样本的误差。这些误差一般都不为零是导致 $\hat{\beta}_1$ 与 β_1 有差异的原因。

利用方程(2.52)的表述，可以证明 OLS 最重要的统计学性质。

定理 2.1（OLS 的无偏性）

利用假定 SLR.1～SLR.4，我们有

$$\mathrm{E}(\hat{\beta}_0)=\beta_0\ 且\ \mathrm{E}(\hat{\beta}_1)=\beta_1\tag{2.53}$$

对 β_0 和 β_1 的任何值都成立。换言之，$\hat{\beta}_0$ 对 β_0 是无偏的，$\hat{\beta}_1$ 对 β_1 是无偏的。

证明 在证明中，期望值是以自变量的样本值为条件的。因为 s_x^2 和 d 是 x 的函数，所以它们在条件中是非随机的。因此，由式（2.53），有

$$\begin{aligned}\mathrm{E}(\hat{\beta}_1)&=\beta_1+\mathrm{E}\Big[(1/s_x^2)\sum_{i=1}^{n}d_iu_i\Big]\\&=\beta_1+(1/s_x^2)\sum_{i=1}^{n}\mathrm{E}(d_iu_i)\\&=\beta_1+(1/s_x^2)\sum_{i=1}^{n}d_i\mathrm{E}(u_i)\\&=\beta_1+(1/s_x^2)\sum_{i=1}^{n}d_i\cdot0=\beta_1\end{aligned}$$

这里已经利用了下面这个事实，即在假定 SLR.2 和 SLR.3 之下，每个 u_i 的期望值（以 $\{x_1,x_2,\cdots,x_n\}$ 为条件）均为零。

现在，对 $\hat{\beta}_0$ 的证明就非常显然了。方程（2.48）对 i 取均值，得到 $\bar{y}=\beta_0+\beta_1\bar{x}+\bar{u}$，将其代入 $\hat{\beta}_0$ 的公式

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=\beta_0+\beta_1\bar{x}+\bar{u}-\hat{\beta}_1\bar{x}=\beta_0+(\beta_1-\hat{\beta}_1)\bar{x}+\bar{u}$$

那么，以 x_i 的值为条件，有

$$\mathrm{E}(\hat{\beta}_0)=\beta_0+\mathrm{E}[(\beta_1-\hat{\beta}_1)\bar{x}]+\mathrm{E}(\bar{u})=\beta_0+\mathrm{E}[(\beta_1-\hat{\beta}_1)]\bar{x}$$

因为根据假定 SLR.2 和 SLR.3，E（$\bar{u}$）＝0。但是我们知道 E（$\hat{\beta}_1$）＝β_1，这就表示 E［（$\hat{\beta}_1-\beta_1$）］＝0。因此，E（$\hat{\beta}_0$）＝β_0。以上两个推导对 β_0 和 β_1 的任何值都有效，由此我们就建立了 OLS 的无偏性。

51 必须记住，无偏性是 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的抽样分布的性质，并没有告诉我们从特定样本中得到的估计值是什么。我们的希望是，如果得到的样本是“典型”的，那么我们的估计值就会“接近于”总体值。不幸的是，我们总有可能得到点估计远离 β_1 的不幸样本，而且我们永远也不可能确知情况是否如此。或许你需要附录 C 来复习无偏估计量的内容，特别是表 C.1 中说明无偏性概念的仿真练习。

如果我们的四个假定中有一个不成立，那么无偏性一般也不成立。这就是说，意识到每一个假定对具体样本所显示的真实性是非常重要的。正如我们讨论过的，如果假定 SLR.4 不成立，就得不到 OLS 的估计值。假定 SLR.1 要求 y 和 x 有线性关系，还要加上一个干扰项，这有可能是不成立的。但我们同样知道，可以通过选择 y 和 x 来建立我们感兴趣的非线性关系。解决方程（2.47）的失效问题要求较高深的方法，而这已经超出了本书的讨论范围。

今后，我们要为时间序列分析放宽关于随机抽样的假定 SLR.2。但是用它来分析横截面会怎样呢？当样本不能代表背后的总体时，随机抽样在横截面上就会不成立。事实上，有些数据集合就是通过有意识地从总体中的不同部分过度取样建立起来的。我们将在第 9 章和第 17 章讨论非随机抽样的问题。

现在着重讨论假定 SLR.3。如果 SLR.3 成立的话，OLS 估计量就是无偏的。同样地，如果 SLR.3 不成立，那么 OLS 估计量一般就是**有偏的**。用以确定偏差的方向和大小的方法是有的，我们将在第 3 章学习它们。

在非实验性数据的简单回归分析中，x 与 u 相关的可能性，正如在 2.1 节给出的几个例子所表明的那样是我们关注的问题。当 u 包含着影响 y 且与 x 相关的因素时，利用简单回归就会导致**谬误相关**。就是说，我们发现 y 和 x 的关系是由既影响着 y 同时又恰巧与 x 相关的不可测因素影响着的。

例 2.12　学生的数学考试成绩和学校的午餐计划

令 *math*10 表示一个高中的十年级学生在一次标准化数学考试中获得通

过的百分比。假令我们的希望是，估计联邦资助的学校午餐计划对学生成绩的影响。我们估计午餐计划对学生成绩有在其他条件不变下的正的影响，即其他条件相同时，如果一个学生太贫穷而不能保证有规律的饮食，就可以获得学校午餐计划资助的资格，他或她的成绩就会得到提高。用 *lnchprg* 表示有资格接受午餐计划的学生的百分比。那么简单回归模型就是

$$math10 = \beta_0 + \beta_1 lnchprg + u \tag{2.54}$$

52 学校和学生的特性中能影响整个学校表现的因素被包含在 u 中。MEAP93.RAW 包括了 408 所密歇根的高中 1992—1993 学年的数据，利用这些数据，我们得到

$$\widehat{math10} = 32.14 - 0.319 lnchprg$$
$$n = 408,\ R^2 = 0.171$$

这个方程表明，如果有资格接受午餐计划的学生增加 10 个百分点，通过数学考试的学生会减少 3.2 个百分点。午餐计划更高的参与率实际上导致了更糟糕的成绩，我们真的能够相信这一点吗？答案几乎是否定的。更好的解释是方程（2.54）中的误差项 u 和 *lnchprg* 是相关的。事实上，u 包含着诸如学校学生的贫穷率这样既影响学生成绩又与午餐计划资格高度相关的因素。像学校质量和资源这样的变量也被包含在 u 中，它们都可能与 *lnchprg* 相关。应该记住，估计值 −0.319 只针对这个特定例子，但它的符号和大小让我们怀疑 u 和 x 可能是相关的，因此该简单回归是有偏误的。

在简单回归模型中，除了遗漏变量，还有其他原因使 x 与 u 相关。因为同样的问题也会出现在多元回归分析中，所以我们要推迟到那时再系统地讨论这个问题。

OLS 估计量的方差

除了知道 $\hat{\beta}_1$ 的抽样分布是以 β_1 为中心（$\hat{\beta}_1$ 是无偏的）的之外，能了解我们预期的 $\hat{\beta}_1$ 究竟距离 β_1 平均有多远，也是非常重要的。在其他问题中间，这个问题让我们得以从所有的估计量中选择最佳的一个，或者至少是由一些无偏估计量组成的广泛的群体。最容易操作的是度量 $\hat{\beta}_1$(和 $\hat{\beta}_0$)分布的分散程度的方差或者它的平方根，即标准差。(见附录 C 中更详细的讨论。)

结果表明，在假定 SLR.1～SLR.4 下，OLS 估计量的方差可以计算出来。然而，这些表达式会比较复杂。有鉴于此，我们增加一个在横截面分析中较为传统的假定。这个假定要求以 x 为条件的不可观测变量 u 的方差是一个常数。这就是**同方差**（homoskedasticity）或“常方差”假定。

假定 SLR.5（同方差性）

$$\mathrm{Var}(u \mid x) = \sigma^2$$

我们必须强调，同方差假定与零条件均值假定即$E(u|x)=0$非常不同。假定SLR.3涉及的是u的**期望值**，但是假定SLR.5关心的是u的**方差**（均以x为条件）。回忆在没有假定SLR.5时所建立的OLS无偏性，可以知道，同方差假定对于表明$\hat{\beta}_0$和$\hat{\beta}_1$的无偏性毫无作用。我们增加假定SLR.5是因为它简化了$\hat{\beta}_0$和$\hat{\beta}_1$的方差运算，而且它还显示了普通最小二乘法的某种有 53 效性质，这一点我们会在第3章看到。如果假设u和x是独立的，那么给定x时u的分布就不依赖于x，因此$E(u|x)=E(u)=0$且$\mathrm{Var}(u|x)=\sigma^2$。但独立对一个假定来说也许太强了。

因为$\mathrm{Var}(u|x)=E(u^2|x)-[E(u|x)]^2$且$E(u|x)=0,\sigma^2=E(u^2|x)$，这意味着$\sigma^2$也是$u^2$的**无条件期望值**。因为$E(u)=0$，所以$\sigma^2=E(u^2)=\mathrm{Var}(u)$。换言之，$\sigma^2$是$u$的**无条件方差**，所以$\sigma^2$经常被叫做**误差方差**（error variance）或干扰方差。σ^2的平方根σ是误差的标准误。较大的σ表示影响y的不可观测因素的分布较为分散。

把假定SLR.3和SLR.5用y的条件均值和条件方差的形式写出来常常会对我们有所帮助：

$$E(y|x)=\beta_0+\beta_1x \tag{2.55}$$

$$\mathrm{Var}(y|x)=\sigma^2 \tag{2.56}$$

换言之，给定x的y的条件期望值与x呈线性关系，但是给定x的y的方差是恒定的。这一关系由图2.8表现出来，图中$\beta_0>0$且$\beta_1>0$。

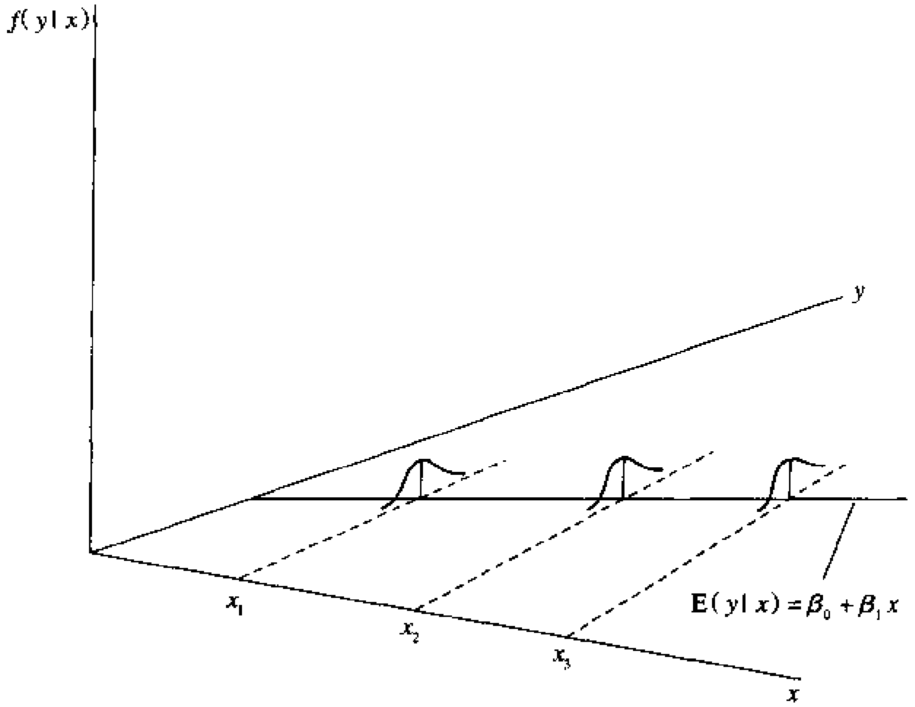


图2.8 同方差下的简单回归分析

当$\mathrm{Var}(u|x)$依赖于x时，误差项就呈现**异方差性**（heteroskedasticity） 54 （或者非恒定方差）。因为$\mathrm{Var}(u|x)=\mathrm{Var}(y|x)$，所以只要$\mathrm{Var}(y|x)$是$x$的函数，异方差就出现了。

例 2.13　工资方程中的异方差

为得到教育对工资在其他条件不变下的影响的无偏估计量，我们必须假设$E(u|educ)=0$，这表示$E(wage|educ)=\beta_0+\beta_1 educ$。如果同样作出同方差假定，那么$\mathrm{Var}(u|educ)=\sigma^2$不依赖于教育水平，这与假定$\mathrm{Var}(wage|educ)=\sigma^2$的效果相同。因此虽然平均工资可随教育水平的提高而增长——这个增长率正是我们描述的兴趣方程所在——但工资相对于它的均值的变异却被假定为对所有的教育水平均不变。这或许不符合实际。因为接受了更多教育的人可能有更广泛的兴趣和更多的就业机会，这就导致了更高的教育会产生更大的工资变异。而教育水平低的人工作机会少，而且只能得到最低限额的工资，这就使得较低的教育水平减少了工资的变异。这一情况通过图 2.9 表现出来。最后，假定 SLR.5 是否成立是一个经验性问题，而且在第 8 章中我们会说明如何检验假定 SLR.5。

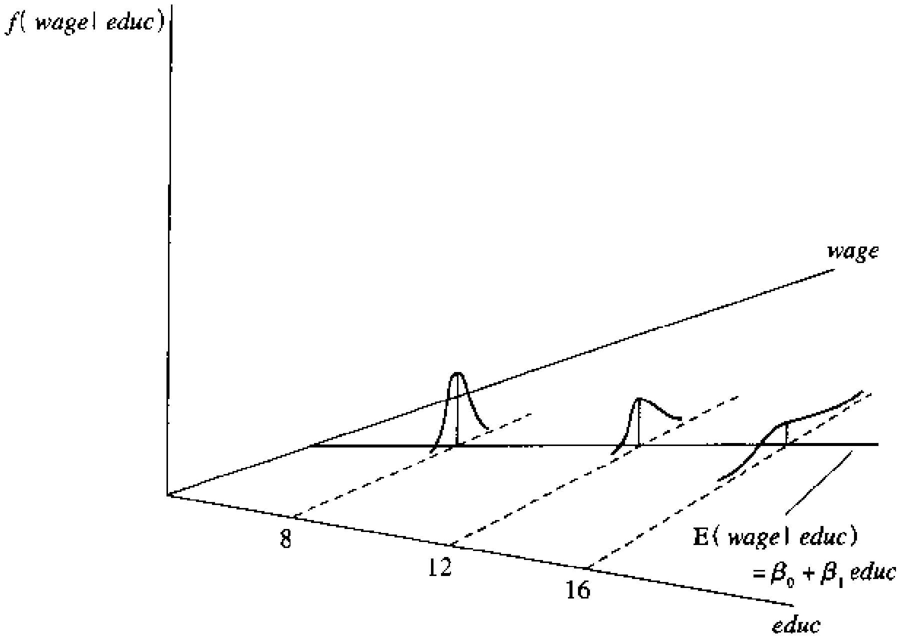


图 2.9　Var(wage|educ)随 educ 增加

55　有了同方差假定，我们就可以证明下面的定理了。

定理 2.2（OLS 估计量的抽样方差）

在假定 SLR.1～SLR.5 下，有

$$\mathrm{Var}(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\sigma^2/s_x^2 \tag{2.57}$$

$$\mathrm{Var}(\hat{\beta}_0)=\frac{\sigma^2 n^{-1}\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2.58}$$

这些结果是以样本值 $\{x_1, x_2\cdots, x_n\}$ 为条件的。

证明 我们只推导 $\mathrm{Var}(\hat{\beta}_1)$ 的公式，把另一个推导留作练习。从方程 (2.52) 开始：$\hat{\beta}_1 = \beta_1 + (1/s_x^2)\sum_{i=1}^{n} d_i u_i$，因为 β_1 只是一个常数，而且以 x_i 为条件，所以 s_x^2 和 $d_i = x_i - \bar{x}$ 同样是非随机的。其次，因为 u_i 是第 i 次观测（通过随机抽样）的独立随机变量，故和的方差就是方差的和。利用这个事实，我们有

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) &= (1/s_x^2)^2 \mathrm{Var}\Big(\sum_{i=1}^{n} d_i u_i\Big)\\ &= (1/s_x^2)^2 \Big(\sum_{i=1}^{n} d_i^2 \mathrm{Var}(u_i)\Big)\\ &= (1/s_x^2)^2 \Big(\sum_{i=1}^{n} d_i^2 \sigma^2\Big)[\text{因为对所有的 } i \text{ 有 } \mathrm{Var}(u_i) = \sigma^2]\\ &= \sigma^2 (1/s_x^2)^2 \Big(\sum_{i=1}^{n} d_i^2\Big) = \sigma^2 (1/s_x^2)^2 s_x^2 = \sigma^2 / s_x^2\end{aligned}$$

这正是我们想要展示的。

式 (2.57) 和式 (2.58) 是简单回归分析的“标准”公式，在异方差出现的时候就会失效。在我们考虑多元回归分析中的置信区间和假设检验的问题时，它们的重要性就表现出来了。

对大多数目的来说，我们关注的都是 $\mathrm{Var}(\hat{\beta}_1)$。这个方差是如何决定于误差方差 σ^2 和 $\{x_1, x_2, \cdots, x_n\}$ 的总变异 s_x^2 的，不难加以概括。首先，误差方差越大，$\mathrm{Var}(\hat{\beta}_1)$ 越大。因为影响 y 的不可观测因素的变异越大，要准确估计 β_1 就越难；另一方面，自变量的变异越大越好，因为当 x_i 的变异性增加时，$\hat{\beta}_1$ 的方差就会减小。这一点也符合直觉，因为自变量的样本分布得
56 越分散，就越容易找出 $\mathrm{E}(y|x)$ 和 x 之间的关系。也就是说，越容易估计出 β_1。如果 x_i 没有什么变化，就难以准确地确定 $\mathrm{E}(y|x)$ 是如何随着 x 变化的。当样本容量扩大时，x_i 的总体变化也增加。因此，较大的样本容量会产生较小的 $\hat{\beta}_1$ 的方差。

问题 2.5

在估计 β_0 的时候，最好是有 $\bar{x} = 0$。在这种情况下 $\mathrm{Var}(\hat{\beta}_0)$ 是多少呢？[提示：对任何样本中的数值，都有 $\sum_{i=1}^{n} x_i^2 \geqslant \sum_{i=1}^{n}(x_i - \bar{x})^2$，当且仅当 $\bar{x} = 0$ 时取等号。]

这一分析表明，如果我们对 $\hat{\beta}_1$ 感兴趣，就可以选择使 x_i 尽可能分散。这在实验性数据中是可能的，但在社会科学中却极少有这么丰富的样本供我们选择，因为通常要通过随机抽样来得到 x_i。但有时候我们还是有机会得到较大的样本的，尽管代价或许会很高。

为了建立置信区间和导出检验统计量，我们需要用到 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的标准误

$\mathrm{sd}(\hat{\beta}_1)$和$\mathrm{sd}(\hat{\beta}_0)$。回忆一下我们就可以知道，这可通过取式（2.57）和式（2.58）中的方差的平方根得到。特别地，我们有$\mathrm{sd}(\hat{\beta}_1)=\sigma/s_x$，其中$\sigma$是$\sigma^2$的平方根，$s_x$是$s_x^2$的平方根。

误差方差的估计

式（2.57）和式（2.58）使我们能够将影响$\mathrm{Var}(\hat{\beta}_1)$和$\mathrm{Var}(\hat{\beta}_0)$的因素分离出来。但除非是在已知$\sigma^2$这样极端罕见的情况下，否则这些公式是不可确知的。不过可以用观测数据去估计σ^2，从而估计出$\mathrm{Var}(\hat{\beta}_1)$和$\mathrm{Var}(\hat{\beta}_0)$。

现在让我们来强调**误差**（或干扰）与**残差**的区别再合适不过了，因为这一区别对建立σ^2的估计量非常关键。方程（2.48）告诉我们如何利用诸如$y_i=\beta_0+\beta_1x_i+u_i$的样本观测值来写出总体模型，这里$u_i$是第$i$次观测的误差。我们还可以像方程（2.32）那样用$y_i$的拟合值和残差把它表示出来：$y_i=\hat{\beta}_0+\hat{\beta}_1x_i+\hat{u}_i$。比较这两个方程，可以看出，误差出现在包含着**总体**参数β_0和β_1的方程中；另一方面，残差在估计方程中与$\hat{\beta}_0$和$\hat{\beta}_1$一起出现。误差是无法观测的，但是残差却可以从数据中计算出来。

我们可以用方程（2.32）和方程（2.48）把残差写成误差的函数：

$$\hat{u}_i=y_i-\hat{\beta}_0-\hat{\beta}_1x_i=(\beta_0+\beta_1x_i+u_i)-\hat{\beta}_0-\hat{\beta}_1x_i$$

或者

$$\hat{u}_i=u_i-(\hat{\beta}_0-\beta_0)-(\hat{\beta}_1-\beta_1)x_i \tag{2.59}$$

尽管$\hat{\beta}_0$的期望值等于β_0，$\hat{\beta}_1$的期望值也等于β_1，$\hat{u}_i$却不等于u_i。但两者之间的差异的**期望值**倒确实为零。

57 只要理解了误差和残差的区别，就可以估计σ^2了。首先，$\sigma^2=\mathrm{E}(u^2)$，所以σ^2的无偏“估计量”是$n^{-1}\sum_{i=1}^{n}u_i^2$。不幸的是，这不是一个真实的估计量，因为我们观察不到误差u_i。但是我们有u_i的估计量，即OLS残差$\hat{u}_i$。如果用OLS残差来代替误差，就有$n^{-1}\sum_{i=1}^{n}\hat{u}_i^2=\mathrm{SSR}/n$。这是一个真实的估计量，因为它给出了关于$x$和$y$的任何数据样本一个可计算的规则。这个估计量的小毛病是它被证明是有偏误的（尽管对大样本来说这个偏差很小）。但是因为从它易于计算出一个无偏估计量，所以我们还是用后者来作为替代。

估计量SSR/n从本质上说是有偏误的，因为它没有考虑必须被OLS残差所满足的两个限制条件。这些限制由OLS的两个一阶条件给出：

$$\sum_{i=1}^{n}\hat{u}_i=0,\quad\sum_{i=1}^{n}x_i\hat{u}_i=0 \tag{2.60}$$

考虑这些限制的一个方法是，如果我们知道残差中的$n-2$个，就能够通过

式（2.60）中的一阶条件所蕴涵的限制得到另外两个残差。因此，OLS残差只有 $n-2$ 个自由度。［与误差的 n 个自由度相对应。如果在式（2.60）中用 u_i 取代 $\hat{u}_i$，这些限制就消失了。］我们将要用到的 σ^2 的无偏估计量对自由度作出了调整：

$$\hat{\sigma}^2 = \frac{1}{(n-2)}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{\mathrm{SSR}}{n-2} \tag{2.61}$$

（这个估计量有时被表示为 s^2，但我们继续使用惯常的方法，在估计量上加“帽”。）

定理 2.3（σ^2 的无偏估计）

在假定 SLR.1～SLR.5 下，我们有

$$\mathrm{E}(\hat{\sigma}^2)=\sigma^2$$

证明 如果把方程(2.59)对所有 i 取平均值，并利用OLS的残差的平均值为零，我们有 $0=\bar{u}-(\hat{\beta}_0-\beta_0)-(\hat{\beta}_1-\beta_1)\bar{x}$；把它从式(2.59)中减去，就得到 $\hat{u}_i=(u_i-\bar{u})-(\hat{\beta}_1-\beta_1)(x_i-\bar{x})$。因此，$\hat{u}_i^2=(u_i-\bar{u})^2+(\hat{\beta}_1-\beta_1)^2(x_i-\bar{x})^2-2(u_i-\bar{u})(\hat{\beta}_1-\beta_1)(x_i-\bar{x})$。将所有 i 次观测相加，得到 $\sum_{i=1}^{n}\hat{u}_i^2=\sum_{i=1}^{n}(u_i-\bar{u})^2+(\hat{\beta}_1-\beta_1)^2\sum_{i=1}^{n}(x_i-\bar{x})^2-2(\hat{\beta}_1-\beta_1)\sum_{i=1}^{n}u_i(x_i-\bar{x})$。现在由附录C可以看到，等式右边第一项的期望值是 $(n-1)\sigma^2$。第二项的期望值只是简单的 σ^2，因为 $\mathrm{E}[(\hat{\beta}_1-\beta_1)^2]=\mathrm{Var}(\hat{\beta}_1)=\sigma^2/s_x^2$。最后，第三项可以写为 $2(\hat{\beta}_1-\beta_1)^2s_x^2$，取期望得到 $2\sigma^2$。把这三项放在一起，得到 $\mathrm{E}(\sum_{i=1}^{n}\hat{u}_i^2)=(n-1)\sigma^2+\sigma^2-2\sigma^2=(n-2)\sigma^2$，因此 $\mathrm{E}[\mathrm{SSR}/(n-2)]=\sigma^2$。

58 如果把 $\hat{\sigma}^2$ 插入方差公式（2.57）和（2.58），就得到 $\mathrm{Var}(\hat{\beta}_1)$ 和 $\mathrm{Var}(\hat{\beta}_0)$ 的无偏估计量。今后要用到 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的标准误的估计量，而这要求我们估计 σ。σ 的自然估计量为

$$\hat{\sigma}=\sqrt{\hat{\sigma}^2} \tag{2.62}$$

我们称其为**回归的标准误**（standard error of the regression，SER）。（$\hat{\sigma}$ 的其他名字有**估计值的标准误**和**均方误的平方根**，但我们不会用到这些。）尽管 $\hat{\sigma}$ 不是 σ 的无偏估计量，可是我们能够证明它是 σ 的一致估计量（见附录C），而且它能够很好地服务于我们的目的。

估计值 $\hat{\sigma}$ 值得关注，因为它是对影响 y 的不可观测因素的标准误的估计；等价地，它估计了把 x 的影响排除之后 y 的标准误。大部分回归的软件包会给出 $\hat{\sigma}$ 的值，还有 R-平方、截距、斜率和其他OLS统计量（用上面列出的名字中选出某一个给出）。现在，我们主要关注的是利用 $\hat{\sigma}$ 来估计 $\hat{\beta}_0$ 和

$\hat{\beta}_1$ 的标准误。因为 $\mathrm{sd}(\hat{\beta}_1)=\sigma/s_x$，$\mathrm{sd}(\hat{\beta}_1)$ 的自然估计量为

$$\mathrm{se}(\hat{\beta}_1)=\frac{\hat{\sigma}}{s_x}=\frac{\hat{\sigma}}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{1/2}}$$

它被称为 **$\hat{\beta}_1$ 的标准误**（standard error of $\hat{\beta}_1$）。当考虑对 y 的不同样本使用 OLS 时，我们要注意将 $\mathrm{se}(\hat{\beta}_1)$ 看做一个随机变量，这是因为 $\hat{\sigma}$ 是随着样本的不同而变化的。对于一个给定的样本，$\mathrm{se}(\hat{\beta}_1)$ 是一个数字，就像当我们用给定的数据计算 $\hat{\beta}_1$ 时那样，它也只是一个数字。

类似地，$\mathrm{se}(\hat{\beta}_0)$ 是通过将 $\mathrm{sd}(\hat{\beta}_0)$ 中的 σ 置换成 $\hat{\sigma}$ 得到的。任何一个估计值的标准误都能让我们了解一个估计量的精确性有多少。标准误在本书中担当着重要角色；从第 4 章开始，我们将要用它们来建立所要研究的任何计量经济程序中的检验统计量和置信区间。

2.6 过原点回归

在少数例子中，我们希望加入下面这个限制：当 $x=0$ 时，y 的期望值为零。有一些特定的关系使得这个限制显得合理。例如，如果收入（x）为零，那么收入税所得（y）也必须为零。另外，有一些问题就把一个最初有非零截距的模型转化成为零截距模型。

形式上，我们选择一个斜率估计量，把它叫做 $\tilde{\beta}_1$，它的曲线为

$$\tilde{y}=\tilde{\beta}_1 x \tag{2.63}$$

β_1 和 y 上面的波浪号用来将这个问题与同时估计截距和斜率的更为常见的问题区别开来。因为曲线（2.63）经过点 $x=0$，$\tilde{y}=0$，所以得到的方程（2.63）又被叫做**过原点回归**（regression through the origin）。为了得到曲线（2.63）中的斜率估计值，仍然需要用到普通最小二乘的技巧。在这个例子中就是使得残差的平方和最小

59

$$\sum_{i=1}^{n}(y_i-\tilde{\beta}_1 x_i)^2 \tag{2.64}$$

通过计算，可以看到，$\tilde{\beta}_1$ 必须满足一阶条件

$$\sum_{i=1}^{n}x_i(y_i-\tilde{\beta}_1 x_i)=0 \tag{2.65}$$

依此可以解出 $\tilde{\beta}_1$：

$$\tilde{\beta}_1=\frac{\sum_{i=1}^{n}x_i y_i}{\sum_{i=1}^{n}x_i^2} \tag{2.66}$$

条件是并非所有的 x_i 都为零，而这一情形我们已经排除。

注意与同时估计截距（而不是把它设置为零）的斜率估计值相比 $\tilde{\beta}_1$ 有什么不同。当且仅当 $x=0$ 时，这两个估计值才是相同的。[见关于 $\tilde{\beta}_1$ 的方程（2.49）。] 在应用研究中，用过原点的回归得到 β_1 的情形并不常见，这也是有道理的，因为如果截距 $\beta_0 \neq 0$，那么 $\tilde{\beta}_1$ 是 β_1 的有偏估计量。习题 2.8 要求你证明它。

▶ 小 结

在这一章中，我们介绍了简单线性回归模型，还讲到了它的基本性质。给定一个随机样本，普通最小二乘法被用来估计总体模型中的斜率和截距参数。我们已经说明了 OLS 回归线的代数性质，包括拟合值和残差的计算，还有如何在给定自变量的变化时预测因变量的变化。在 2.4 节中，我们讨论了在实践中极具重要性的两个问题：(1) 当我们改变因变量或者自变量的测量单位时，OLS 估计值的变化；(2) 利用自然对数来考虑常弹性和常半弹性模型。

在 2.5 节中，我们证明，在 SLR.1－SLR.4 这四个假定之下，OLS 估计量是无偏的。关键的假定是对任何给定的自变量 x 的值，误差项 u 的均值为零。不幸的是，我们有理由相信，在许多简单回归的社会科学的应用中，它是错误的，因为 u 中被忽略的因素经常都是与 x 相关的。当我们加入一个假定，以说明给定 x 的误差的方差是一个常量时，我们就得到了 OLS 估计量的抽样方差的简单公式。正如我们所见，误差方差增加，斜率估计量 $\hat{\beta}_1$ 的方差也增加，而当自变量的样本变化增加时，$\hat{\beta}_1$ 的方差就减少。我们还推导出 $\sigma^2 = \mathrm{Var}(u)$ 的无偏估计量。

在 2.6 节中，简单讨论了过原点回归，我们由零截距的假定得到斜率的估计量。有时它是有用的，但在应用研究中用得不多。

60 还有许多工作尚未完成。例如，我们仍然不知道如何检验关于总体参数 β_0 和 β_1 的假设。因此尽管我们知道在假定 SLR.1－SLR.4 之下 OLS 对总体参数是无偏的，但还是无法作出关于总体的推断。其他一些问题，例如相对于其他可能估计程序的 OLS 的有效性问题，也被忽略了。

置信区间、假设检验和有效性的问题对多元回归分析也是非常关键的。因为我们建立置信区间和检验统计的方法与多元回归分析非常相似——又因为简单回归是多元回归的特例——所以，现在最好进入多元回归的讨论，而且它比简单回归的应用也要广泛一些。第 2 章的目的是启发您在非常简单的背景之下思考计量经济学分析中出现的问题。

关键术语

判定系数
常弹性模型
控制变量
协变量
自由度
因变量
弹性
误差项（干扰项）
误差方差
解释平方和
被解释变量
解释变量
一阶条件
拟合值
异方差性
同方差性
自变量
截距参数
普通最小二乘法
OLS 回归线
总体回归函数
被预测变量
预测元变量
回归子
过原点回归
回归元
残差
残差平方和
响应变量
R-平方
样本回归函数
半弹性
简单线性回归模型
斜率参数
$\hat{\beta}_1$ 的标准误
回归的标准误
残差的平方和
总平方和
零条件均值假定

习　题

2.1　令 *kids* 表示一名妇女生育的孩子数目，*educ* 表示该妇女接受过教育的年数。生育率对教育年数的简单回归模型为

$$kids = \beta_0 + \beta_1 educ + u$$

u 是观测不到的误差。

(i) u 中包含什么样的因素？它们可能与教育水平相关吗？

(ii) 简单回归分析能够揭示教育对生育率在其他条件不变下的影响吗？请解释。

61　2.2　在简单线性回归模型 $y = \beta_0 + \beta_1 x + u$ 中，假定 $\mathrm{E}(u) \neq 0$。令 $\alpha_0 = \mathrm{E}(u)$，证明：这个模型总可以改写为另一种形式，斜率与原来的相同，但

是截距和误差有所不同，并且新的误差有一个零期望值。

2.3　下表包括 8 个学生的 ACT 分数和 GPA（平均积点分）。平均积点分是以四分制计算的且已被四舍五入到小数点后一位。

学生	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

（i）利用 OLS 估计 GPA 和 ACT 的关系；就是说，求下列方程式中的截距和斜率的估计值

$$\widehat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT$$

评价这个关系式。这里的截距有没有适用的解释？请说明之。如果 ACT 分数增加 5 分，GPA 预料会升高多少？

（ii）计算每次观测的拟合值和残差，并证明残差的和（近似）为零。

（iii）当 $ACT = 20$ 时，GPA 的预测值为多少？

（iv）对这 8 个学生来说，GPA 的变异有多少是由 ACT 解释了的？试说明之。

2.4　数据集 BWGHT.RAW 包含有美国妇女生育率的数据。两个变量分别是因变量：婴儿出生时体重的盎司数（$bwght$）和解释变量：母亲在怀孕期间平均每天的吸烟数（$cigs$）。下面这个简单回归是用 $n = 1\,388$次出生数据进行估计的：

62

$$\widehat{bwght} = 119.77 - 0.514 cigs$$

（i）当 $cigs = 0$ 时，婴儿的预测出生体重为多少？当 $cigs = 20$（每天一包）时呢？评价其差别。

（ii）这个简单回归能够得到婴儿出生体重和母亲吸烟习惯之间的关系吗？解释之。

2.5　下式为线性消费函数

$$\widehat{cons} = \hat{\beta}_0 + \hat{\beta}_1 inc$$

（估计的）收入的边际消费倾向（MPC），就是斜率 $\hat{\beta}_1$，而**平均消费倾向**（APC）则是 $\widehat{cons}/inc = \hat{\beta}_0/inc + \hat{\beta}_1$。利用对 100 个家庭年收入和消费的观测（均用美元计算），得到

$$\widehat{cons} = -124.84 + 0.853inc$$

$$n = 100,\ R^2 = 0.692$$

(i) 解释式中的截距并评价它的符号和大小。

(ii) 当家庭收入为30 000时，预测消费为多少？

(iii) 以 inc 为 x 轴，画出估计的 MPC 和 APC 的图。

2.6 利用基尔和麦克莱恩（Kiel and McClain，1995）关于 1988 年安德沃 MA 出售房屋的数据，下面的方程表示房屋价格（$price$）和距离一个新近修建的垃圾焚化炉（$dist$）的远近的关系：

$$\log(\widehat{price}) = 9.40 + 0.312\log(dist)$$

$$n = 135,\ R^2 = 0.162$$

(i) 解释 $\log(dist)$ 的系数。它的符号是你所预期的吗？

(ii) 你认为简单回归给出了 $price$ 对 $dist$ 在其他条件不变下的弹性的无偏估计量吗？（考虑一个城市决定放置焚化炉的地点的决策。）

(iii) 还有其他哪些因素影响房价？这些因素会与焚化炉的距离相关吗？

2.7 考虑储蓄函数

$$sav = \beta_0 + \beta_1 inc + u,\ u = \sqrt{inc \cdot e}$$

式中，e 为随机变量，且有 $\mathrm{E}(e) = 0$ 和 $\mathrm{Var}(e) = \sigma_e^2$，假设 e 与 inc 不相关。

(i) 证明：若 $\mathrm{E}(u \mid inc) = 0$，则满足零条件均值的关键假设（假定 SLR.3）。［提示：如果 e 独立于 inc，那么 $\mathrm{E}(e \mid inc) = \mathrm{E}(e)$。］

63 (ii) 证明：若 $\mathrm{Var}(u \mid inc) = \sigma_e^2 inc$，则不满足同方差假定 SLR.5。特别地，$sav$ 的方差随着 inc 而增加。［提示：如果 e 和 inc 是不相关的，就有 $\mathrm{Var}(e \mid inc) = \mathrm{Var}(e)$。］

(iii) 储蓄的方差随着家庭收入而增加，作出支持该结论的讨论。

2.8 在假定 SLR.1～SLR.4 之下考虑标准简单回归模型 $y = \beta_0 + \beta_1 x + u$。通常的 OLS 估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_0$ 对它们各自的总体参数是无偏的。令 $\tilde{\beta}_1$ 表示通过截距假定为零的 β_1 的估计量（见 2.6 节）。

(i) 用 x_i，β_0 和 β_0 表示 $\tilde{\beta}_1$。证明当总体截距（β_0）为零时，$\tilde{\beta}_1$ 对 β_1 是无偏的。有没有其他的情况使得 $\tilde{\beta}_1$ 也是无偏的？

(ii) 求 $\tilde{\beta}_1$ 的方差。［提示：方差不依赖于 β_0。］

(iii) 证明 $\mathrm{Var}(\tilde{\beta}_1) \leqslant \mathrm{Var}(\hat{\beta}_1)$。［提示：对任何数据样本，$\sum_{i=1}^{n} x_i^2 \geqslant \sum_{i=1}^{n}(x_i - \bar{x})^2$，该式严格不等，除非 $\bar{x} = 0$。］

(iv) 讨论当我们要从 $\hat{\beta}_1$ 和 $\tilde{\beta}_1$ 二者中作出选择时，偏误和方差的替代关系。

2.9 (i) 令 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 为 y_i 对 x_i 的回归中的截距和斜率，进行 n 次观测。令 c_1 和 c_2 为常数且 $c_2 \neq 0$。令 $\tilde{\beta}_0$ 和 $\tilde{\beta}_1$ 为 $c_1 y_i$ 对 $c_2 x_i$ 的回归的截距和斜率。证明 $\tilde{\beta}_1 = (c_1/c_2)\hat{\beta}_1$ 且 $\tilde{\beta}_0 = c_1\hat{\beta}_0$，从而证明 2.4 节中关于测量单位的

结论。[提示：为得到 $\tilde{\beta}_1$，把改变了度量单位的 x 和 y 代入式（2.19）。然后用式（2.17）求 $\tilde{\beta}_0$，勿忘把经过度量变换的 x 和 y 以及正确的斜率代入其中。]

(ii) 现在令 $\tilde{\beta}_0$ 和 $\tilde{\beta}_1$ 得自 (c_1+y_i) 对 (c_2+x_i) 的回归（对 c_2 和 c_2 不加任何限制），证明 $\tilde{\beta}_1=\hat{\beta}_1$ 且 $\tilde{\beta}_0=\hat{\beta}_0+c_1-c_2\hat{\beta}_1$。

计算机习题

2.10　401K.RAW 中的数据是帕普克（Papke，1995）所分析的数据的一个子集，帕普克是为了研究 401k 养老金计划的参与率和该计划的慷慨程度之间的关系。变量 *prate* 是有活动账户的合格工人的百分比，也就是我们要解释的变量。慷慨程度的度量是计划的匹配率 *mrate*。这个变量给出了当一个工人为公司贡献了 1 美元时公司为该工人的计划捐献的平均数。例如，如果 $mrate=0.50$，那么工人每贡献 1 美元，公司就配给 50 美分。

(i) 求出计划样本中的平均参与率和平均匹配率。

(ii) 现在估计下面这个简单回归方程

$$\widehat{prate}=\hat{\beta}_0+\hat{\beta}_1 mrate$$

报告你的结果以及样本容量和 R-平方。

(iii) 解释你的方程中的截距，解释 *mrate* 的系数。

(iv) 当 $mrate=3.5$ 时，求出 *prate* 的预测值。这是一个合理的预测吗？解释这里出现的情况。

64 (v) *prate* 的变异有多少是由 *mrate* 解释了的？你认为这是一个足够大的量吗？

2.11　数据集 CEOSAL2.RAW 包含了美国一些公司的首席执行官的信息，变量 *salary* 是以千美元计的年薪，*ceoten* 是已担任公司 CEO 的年数。

(i) 求出样本中的平均年薪和平均任期。

(ii) 有多少位 CEO 尚处于担任 CEO 的第一年（就是说 $ceoten=0$）？最长的 CEO 任期是多少？

(iii) 估计简单回归模型

$$\log(salary)=\beta_0+\beta_1 ceoten+u$$

用通常的形式报告你的结果。多担任一年 CEO，年薪的预测百分比（大约）增加多少？

2.12　利用比德尔和哈默什（Biddle and Hamermesh，1990）中的 SLEEP75.RAW 数据来研究在每周用于睡眠的时间和用于有报酬的工作的时间之间是否有一种替代关系。我们可以用它们中的任何一个作为因变量。为具体起见，让我们来估计模型

$$sleep=\beta_0+\beta_1 totwrk+u$$

式中，*sleep* 为每周用于晚上睡眠的分钟数；*totwrk* 为一周中工作的总分钟数。

(i) 用方程的形式，连同观测的次数和 R^2 报告你的结果。该方程中的截距表示什么？

(ii) 如果 *totwrk* 增加 2 小时，*sleep* 估计要减少多少？你觉得这是一个很大的效应吗？

2.13 利用 WAGE2.RAW 中的数据估计一个简单回归以便用智商得分数（IQ）来解释月工资。

(i) 求出样本中的平均工资和平均 IQ。IQ 的标准差是多少？（总体的 IQ 已标准化为平均值是 100，标准差是 15。）

(ii) 估计一个简单回归模型，要求其中任何一单位的 IQ 变化导致的美元工资变化都是相同的。利用这个模型计算 IQ 增加 15 点时工资的预期变化。IQ 能够解释大部分的工资变异吗？

(iii) 现在再估计一个模型，要求其中 IQ 增长一个点对工资的影响的百分比是相同的。如果 IQ 增加 15 点，预期工资增加的百分比大约是多少？

2.14 在化学工业企业的总体中，令 *rd* 表示用于研究和发展的年支出，令 *sales* 表示年销售额（都用百万美元计算）。

(i) 写一个模型（不是估计的方程式），可以表示 *rd* 和 *sales* 之间有常弹性的。哪一个参数代表弹性？

(ii) 再用 RDCHEM.RAW 中的数据来估计模型。用通常的形式写出估计的方程。*rd* 关于 *sales* 的估计弹性是多少？用文字解释弹性的含义。

附录 2A

65 如 2.2 节所断言的，我们来证明 OLS 估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 确实能使残差的平方和最小。形式上看，问题是要描述下列最小化问题的解 $\hat{\beta}_0$ 和 $\hat{\beta}_0$ 的特征

$$\min_{b_0,b_1}\sum_{i=1}^{n}(y_i-b_0-b_1x_i)^2$$

式中，b_0 和 b_1 是为最优问题而虚拟的量。为简单起见，把它叫做函数 $Q(b_0, b_1)$。由多元微积分的基本结果可知（见附录 A），$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 作为最小化问题的解的必要条件是，在 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 取值时 $Q(b_0, b_1)$ 关于 b_0 和 b_1 的偏导数必须为零：$\partial Q(\hat{\beta}_0,\hat{\beta}_1)/\partial b_0=0$ 且 $\partial Q(\hat{\beta}_0,\hat{\beta}_1)/\partial b_1=0$。利用微积分的链导法则，这两个方程可写为

$$-2\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)=0$$

$$-2\sum_{i=1}^{n}x_i(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)=0$$

这两个方程式不外是式（2.14）和式（2.15）乘以 $-2n$ 罢了，因此它们的解也同样是 $\hat{\beta}_0$ 和 $\hat{\beta}_1$。

我们怎么知道真的使残差的平方和最小化了呢？毕竟一阶条件是必要但不充分条件。证明我们已得到最小的残差平方和的一种方法是对任何 b_0 和 b_1 写出

$$\begin{aligned}
Q(b_0,b_1) &= \sum_{i=1}^{n}[y_i-\hat{\beta}_0-\hat{\beta}_1x_i+(\hat{\beta}_0-b_0)+(\hat{\beta}_1-b_1)x_i]^2\\
&= \sum_{i=1}^{n}[\hat{u}_i+(\hat{\beta}_0-b_0)+(\hat{\beta}_1-b_1)x_i]^2\\
&= \sum_{i=1}^{n}\hat{u}_i^2+n(\hat{\beta}_0-b_0)^2+(\hat{\beta}_1-b_1)^2\sum_{i=1}^{n}x_i^2\\
&\quad +2(\hat{\beta}_0-b_0)(\hat{\beta}_1-b_1)\sum_{i=1}^{n}x_i
\end{aligned}$$

这里我们用到了方程(2.30)和方程(2.31)。残差的平方和不依赖于 b_0 或 b_1，而最后三项的和可以写为

$$\sum_{i=1}^{n}[(\hat{\beta}_0-b_0)+(\hat{\beta}_1-b_1)x_i]^2$$

这可通过简单的代数运算加以证实。因为这是平方和的形式，所以它最小也不小于零。因此，当 $b_0=\hat{\beta}_0$ 且 $b_1=\hat{\beta}_1$ 时其值最小。

第 3 章　多元回归分析：估计

66 我们在第 2 章了解到如何使用简单的回归分析，把因变量 y 解释成一个自变量 x 的函数。经验研究中使用简单回归分析的主要缺陷是，它很难得到 x 在其他条件不变情况下对 y 的影响：关键假定 SLR.3（所有其他影响 y 的因素都与 x 不相关）通常都不现实。

由于**多元回归分析**（multiple regression analysis）允许我们**明确地**控制许多其他也同时影响因变量的因素，所以它更适合于其他条件不变情况下的分析。在我们必须使用非实验数据的情况下，这对检验经济理论和评价经济政策都很重要。由于多元回归模型能够容纳许多可能相关的解释变量，所以在简单回归分析可能误导的情况下，我们可以寄希望于多元回归模型来推断因果关系。

很自然，如果在模型中多增加一些有助于解释 y 的因素，那么，y 的变动就能更多地得到解释。因此，多元回归分析可用于建立更好的因变量预测模型。

多元回归分析的另外一个优点是，它可以用以添加相当一般化的函数关系。在简单的回归模型中，方程中只能出现单一解释变量的一个函数。如我们将看到的那样，多元回归模型的灵活性则大得多。

3.1 节规范地介绍了多元回归模型，并进一步讨论了多元回归与简单回归相比的优势。3.2 节说明了在多元回归模型中如何使用普通最小二乘法来

估计参数。在 3.3 节、3.4 节和 3.5 节，我们描述了 OLS 估计量的各种统计性质，包括无偏性和有效性。

多元回归模型仍是经济学和其他社会科学进行经验分析时使用得最广泛的一个工具。同样，在估计多元回归模型的参数时，普通最小二乘法也很受欢迎。

3.1 使用多元回归的动因

含有两个自变量的模型

我们先用几个简单的例子来说明，如何用多元回归分析来解决简单回归所不能解决的问题。

67 第一个例子是对第 2 章为得到教育对小时工资之影响而引入的工资方程的简单变形：

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u \tag{3.1}$$

式中，*exper* 为在劳动市场上以年计的工作经历。于是工资 *wage* 由受教育水平和工作经历这两个解释变量或自变量及那些观测不到的其他因素来决定。我们首要感兴趣的是，在保持所有其他影响工资的因素不变的情况下，*educ* 对 *wage* 的影响，即我们只对参数 β_1 感兴趣。

与仅联系 *wage* 和 *educ* 的简单回归分析相比，方程（3.1）有效地把 *exper* 从误差项中取出并把它明确地放到方程之中。由于 *exper* 出现在方程中，所以其系数 β_2 度量了 *exper* 在其他条件不变情况下对工资的影响，这本来也有一定的意义。

无足为奇，就像在简单回归中一样，我们将不得不对式（3.1）中的 u 如何与自变量 *educ* 和 *exper* 相关作出假定。但像在 3.2 节中将看到的那样，有一点我们充满信心：因为式（3.1）中明确地包含了工作经历，所以我们就能在保持工作经历不变的情况下，度量教育对工资的影响。如果将工作经历放到误差项的简单回归分析中，就不得不假定工作经历与受教育水平无关，而这又是一个多么脆弱的假定。

第二个例子考虑的问题是，解释在高中阶段对每个学生的平均开支（*expend*）对平均标准化考试成绩（*avgscore*）的影响。假设平均考试成绩取决于学校基金、平均家庭收入（*avginc*）及其他不可观测因素：

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u \tag{3.2}$$

出于政策目的，所关心的系数是 *expend* 在其他条件不变情况下对 *avgscore*

的影响 β_1。通过在模型中明确包括 *avginc*，我们就能控制其对 *avgscore* 的影响。由于平均家庭收入与每个学生的开支趋于相关：开支水平通常由财产税和当地的收入税决定，所以加入这个变量可能很重要。在简单回归分析中，*avginc* 被包括在误差项中，而 *avginc* 与 *expend* 可能相关，从而导致在两变量模型中对 β_1 的 OLS 估计有偏误。

在前面两个相似的例子中，我们已经说明，除主要关心的变量［方程（3.1）中是 *educ*，而方程（3.2）中则是 *expend*］外，如何把其他可观测因素也包括在回归模型中。一般地，我们可以把含有两个自变量的模型写为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \qquad (3.3)$$

式中，β_0 为截距；β_1 度量了在其他条件不变情况下 y 相对 x_1 的变化；β_2 则度量了在其他条件不变情况下 y 相对 x_2 的变化。

多元回归分析对推广变量之间的函数关系也有帮助。例如，假设家庭消费（*cons*）是家庭收入（*inc*）的一个二次函数：

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u \qquad (3.4)$$

68 式中，u 包括了影响消费的其他因素。在这个模型中，消费只取决于收入这一个观测变量，所以看上去一个简单的回归分析就可以对付。但简单回归不能处理这个模型，因为它包括了收入的两个函数 *inc* 和 inc^2（因此就有三个参数 β_0，β_1 和 β_2）。尽管如此，通过令 $x_1 = inc$ 和 $x_2 = inc^2$，消费函数还是可以很容易地写成一个含两个自变量的回归模型。

机械地看，用普通最小二乘法（在 3.2 节介绍过）去估计像式（3.1）和式（3.4）那样不同的方程，应该没有什么差别。每个方程都可以写成像式（3.3）那样的方程，这就是与计算结果有关系的全部。但一个重要的差别在于人们对参数的**解释**。在方程（3.1）中，β_1 是 *educ* 在其他条件不变情况下对 *wage* 的影响，而方程（3.4）中的参数 β_1 则没有这样的解释。换句话说，度量 *inc* 在保持 inc^2 不变的情况下对 *cons* 的影响毫无意义，因为如果 *inc* 变化，则 inc^2 也一定会变化！相反，相对收入变化的消费变化——即边际消费倾向——可近似为

$$\frac{\Delta cons}{\Delta inc} \approx \beta_1 + 2\beta_2 inc$$

对于推导这个方程所需要的微积分，可参见附录 A。换句话说，收入对消费的边际效应取决于 β_2，β_1 和收入水平。这个例子表明，在任何一个特定应用中，对自变量的定义都是至关重要的。但对多元回归的理论发展来说，我们可以对这种细枝末节不甚清楚。在第 6 章，我们将更完备地研究这种例子。

在含有两个自变量的模型中，u 与 x_1 和 x_2 如何相关的关键假定是

$$\mathrm{E}(u \mid x_1, x_2) = 0 \qquad (3.5)$$

对条件（3.5）的解释，与对简单回归分析的假定 SLR.3 的解释相似。它意

味着对总体中 x_1 和 x_2 的任何值，非观测因素的平均都等于零。至于简单回归，假定的重要部分是，u 的期望值对所有 x_1 和 x_2 的组合都是相同的；只要模型中包括了截距 β_0 这一项，根本就不需要上述相同的期望值为零这个假定（参见 2.1 节）。

如何解释前面例子中条件均值为零的假定呢？在方程（3.1）中，这个假定是$\mathrm{E}(u \mid educ, exper)=0$。这意味着影响 $wage$ 的其他因素都与 $educ$ 和 $exper$ 无关。因此，如果认为天生能力是 u 的一部分，那我们就需要对工人总体中受教育和工作经历的各种组合，其平均能力水平都相同。这可能正确也可能不正确，但如在 3.3 节中将看到的那样，正是为了判断普通最小二乘法是否导致无偏估计量而需要知道的问题。

度量学生表现［方程（3.2）］的例子类似于工资方程。其零条件均值的假定为$\mathrm{E}(u \mid expend, avginc)=0$，它意味着影响学生考试成绩的因素——学校或学生的个人特征——总体上与学生的平均开支和平均家庭收入无关。

在应用到式（3.4）中的二次消费函数时，对零条件均值假定的解释则略有不同。直接照写，方程（3.5）就变成了$\mathrm{E}(u \mid inc, inc^2)=0$。因为一旦知道了 inc，那就会知道 inc^2，所以在预期表达式中包括 inc^2 项是多此一举：$\mathrm{E}(u \mid inc, inc^2)=0$ 等价于$\mathrm{E}(u \mid inc)=0$。虽然在表述这个假定时让 inc^2 和 inc 一起出现在预期项中并没有错，但 $\mathrm{E}(u \mid inc)=0$ 更简明扼要。

问题 3.1

69

用定罪概率（$prbconv$）和宣判监禁的增均时间长度（$avgsen$）来解释城市谋杀率（$murdrate$）的一个简单模型是

$$murdrate=\beta_0+\beta_1 prbconv+\beta_2 avgsen+u$$

u 中包含了一些什么因素？你认为关键假定（3.5）有可能成立吗？

含有 K 个自变量的模型

一旦开始多元回归，就没有必要局限于两个自变量。多元回归分析允许多个可观测因素影响 y。在上述工资的例子中，还可以包括在职培训的数量、现任工作的任期、个人能力的某种度量，甚至是兄弟姐妹的个数或母亲受教育程度等人口变量。在学校基金的例子中，额外的变量可能包括对教师质量和学校规模的某种度量。

一般的**多元线性回归模型**（multiple linear regression model，也称为多元回归模型）在总体中可以写成

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\cdots+\beta_k x_k+u \tag{3.6}$$

式中，β_0 为**截距**（intercept）；β_1 为与 x_1 相联系的参数；β_2 为与 x_2 相联系的参数，等等。由于有 k 个自变量和一个截距项，所以方程（3.6）包含了 $k+1$ 个（未知的）总体参数。为了表达上的简便，我们有时把这种不同于

截距的参数称为**斜率参数**，尽管它们并不一定表示斜率。（如方程（3.4），其中 β_1 和 β_2 本身都不是斜率，但它们一起决定了消费与收入之间关系的斜率。）

多元回归的术语类似于简单回归的术语，我们把它们列在表 3.1 中。恰如简单回归中一样，变量 u 表示**误差项**或**干扰项**（disturbance），它包括除 x_1，x_2，…，x_k 之外仍影响 y 的一些因素。无论在我们的模型中包含了多少个解释变量，总有一些因素无法包括进来，而所有这些因素就包括在 u 中。

70 **表 3.1** **多元回归的术语**

y	$x_1, x_2, \cdots, x_k$
因变量	自变量
被解释变量	解释变量
响应变量	控制变量
被预测变量	预测元变量
回归子	回归元

在使用一般多元回归模型时，还必须知道如何解释这些参数。我们现在及在以后的一些章节中将进行大量实践，但此刻联想一些已经知道的东西很有好处。设想 CEO 的薪水（*salary*）与企业的销售量和 CEO 在这个企业的任期相关：

$$\log(salary) = \beta_0 + \beta_1\log(sales) + \beta_2\, ceoten + \beta_3\, ceoten^2 + u \qquad (3.7)$$

通过定义 $y = \log(salary)$，$x_1 = \log(sales)$，$x_2 = ceoten$ 和 $x_3 = ceoten^2$，就得到一个多元回归模型（$k = 3$）。正如我们在第 2 章中所了解的那样，参数 β_1 是（其他条件不变情况下）薪水对销售量的**弹性**。如果 $\beta_3 = 0$，那么在其他条件不变情况下，$100\beta_2$ 就表示 *ceoten* 增加一年导致 *salary* 提高的百分数。当 $\beta_3 \neq 0$ 时，*ceoten* 对 *salary* 的影响则复杂一些。我们把对一个含有二次项的一般模型的详尽处理留待第 6 章讨论。

方程（3.7）就多元回归分析给了我们一个重要的提醒。多元线性回归模型中的“线性”一词，意味着方程（3.6）是其诸参数 β_j 的一个线性函数。方程（3.7）是多元回归模型的一个例子，尽管它线性于 β_j，但 *salary* 与变量 *sales* 和 *ceoten* 之间的关系却是非线性的。多元线性回归的许多运用中都涉及主要变量之间的非线性关系。

一般多元回归模型的关键假定，用条件预期的形式可以很容易地表示为

$$\mathrm{E}(u \mid x_1, x_2, \cdots, x_k) = 0 \qquad (3.8)$$

起码，方程（3.8）要求不可观测的误差项中所有的因素都与解释变量无关。它还意味着，我们已经正确地解释了被解释变量和解释变量之间的函数关系。任何一个导致 u 与某个自变量相关的问题，都会导致式（3.8）不成

立。在3.3节中，我们将证明，假定条件式（3.8）表明OLS是无偏的，而
如果方程中省略了一个关键变量，所得到的结论便会产生偏误。在第15章
71 和第16章中，我们将研究导致式（3.8）不成立的其他原因，并说明在式
（3.8）不能成立的情况下，我们能做些什么。

3.2 普通最小二乘法的操作和解释

现在，让我们总结一下，将普通最小二乘法用于一个特定的数据集时，在计算和代数上会有些什么特征。我们还要讨论如何解释所估计的方程。

如何得到OLS估计值

我们先考虑对含有两个自变量模型的估计。被估计的OLS方程在形式上与简单回归情况下的方程相似：

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2 \tag{3.9}$$

式中，$\hat{\beta}_0$为β_0的估计值；$\hat{\beta}_1$为β_1的估计值；而$\hat{\beta}_2$为β_2的估计值。但如何得到$\hat{\beta}_0$，$\hat{\beta}_1$和$\hat{\beta}_2$呢？**普通最小二乘法**选择能最小化残差平方和的估计值。即给定有关y，x_1和x_2的n个观测$\{(x_{i1}, x_{i2}, y_i): i=1, 2, \cdots, n\}$，同时选择的$\hat{\beta}_0$，$\hat{\beta}_1$和$\hat{\beta}_2$要使下式尽可能小：

$$\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1}-\hat{\beta}_2x_{i2})^2 \tag{3.10}$$

要理解OLS在做什么，重要的是掌握式（3.10）中自变量下标的含义，这里每个自变量都有两个下标，i后面跟着1或2。下标i表示观测序号。于是，式（3.10）是对从1到n所有的观测求和。第二个下标只是区别不同自变量的方法。在*wage*与*educ*和*exper*相关的例子中，$x_{i1}=educ_i$是样本中第i个人的受教育水平，$x_{i2}=exper_i$是第i个人的工作经历。方程（3.10）中的残差平方和为$\sum_{i=1}^{n}(wage_i-\hat{\beta}_0-\hat{\beta}_1educ_i-\hat{\beta}_2exper_i)^2$。以后角标$i$专用于表示观测序号。如果我们写出$x_{ij}$，那它就表示对第$j$个自变量的第$i$个观测值。（有些作者喜欢交换观测序号和变量序号的位置，所以x_{1i}就成了对第一个变量的第i个观测值。但这只是使用记号习惯的不同。）

在含有k个自变量的一般情形中，我们想找到方程

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2+\cdots+\hat{\beta}_kx_k \tag{3.11}$$

中的估计值$\hat{\beta}_0$，$\hat{\beta}_1$和$\hat{\beta}_k$。所选择的$k+1$个OLS估计值最小化了残差平方和

72

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})^2 \tag{3.12}$$

这个最小化问题可通过使用多元微积分求解（参见附录3A）。这样就得到 $\hat{\beta}_0$，$\hat{\beta}_1$ 和 $\hat{\beta}_k$ 这 $k+1$ 个未知变量的 $k+1$ 个线性方程：

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0\\
&\sum_{i=1}^{n}x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0\\
&\sum_{i=1}^{n}x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0\\
&\cdots\cdots\cdots\cdots\\
&\sum_{i=1}^{n}x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0
\end{aligned} \tag{3.13}$$

这个方程组通常被称为**OLS一阶条件**。像在2.2节中的简单回归模型中一样，OLS一阶条件也可以通过矩法得到：在假定条件（3.8）下，$\mathrm{E}(u)=0$，$\mathrm{E}(x_j u)=0(j=1,2,\cdots,k)$。式（3.13）中的方程就是这些总体矩在样本中的对应样本矩。

即便只是中等大小，通过手算来求解方程（3.13）也是十分繁重的任务。不过，借助现代的计算机和统计与计量软件，对较大的 n 和 k，也能很快解出这些方程。

只有一个小小的警告：我们必须假定（3.13）中的方程只能得到 $\hat{\beta}_j$ 的惟一解。目前，我们只能这样假定，因为这是规范设定模型的常见情形。在3.3节，我们将阐述存在惟一的OLS估计值所需要的假设（参见假定MLR.4）。

如在简单回归分析中那样，方程（3.11）被称为**OLS回归线或样本回归方程**。我们把 $\hat{\beta}_0$ 称为**OLS截距估计值**（OLS intercept estimate），而把 $\hat{\beta}_1$，$\hat{\beta}_2$，…，$\hat{\beta}_k$ 称为**OLS斜率估计值**（OLS slope estimate）（与自变量 $x_1, x_2, \cdots, x_k$ 对应）。

为了表明已经进行了一个OLS回归分析，要么将方程（3.13）中的 y 和 x_1，x_2，…，x_k 都用其变量名称取代（如 *wage*，*educ* 和 *exper* 等），或者说"将 y 对 x_1，x_2，…，x_k 进行了一个OLS回归"或"将 y 对 x_1，x_2，…，x_k 进行回归"。这是指出使用普通最小二乘法而得到OLS方程（3.13）的简单说法。除非另有明确说明，否则都把截距与斜率一起估计。

对OLS回归方程的解释

比在计算 $\hat{\beta}_j$ 的背后存在的所有细节都重要的是，对所估计的方程进行**解释**。我们从含有两个自变量的情况开始：

73

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{3.14}$$

方程（3.14）中的截距项 $\hat{\beta}_0$ 是 y 在 $x_1=0$ 和 $x_2=0$ 情况下的预测值。虽然有时令 x_1 和 x_2 都等于零是一个有意义的情况，但在多数情况下都没有什么意义。不仅如此，如方程（3.14）所表明的那样，为了从 OLS 回归中得到 y 的预测值，总是需要截距项。

估计值和局部效应（partial effect）**或其他情况不变效应的解释。**从方程（3.14）中得到

$$\Delta\hat{y}=\hat{\beta}_1\Delta x_1+\hat{\beta}_2\Delta x_2$$

所以能在给定 x_1 和 x_2 的变化的情况下，预测 y 的变化。（注意，截距项怎么会与 y 的变化没有关系呢！）特别是当 x_2 固定，因而 $\Delta x_2=0$ 时，于是，

$$\Delta\hat{y}=\hat{\beta}_1\Delta x_1$$

关键是，通过把 x_2 包含在模型中，我们所得到的 x_1 的系数，可解释为在其他条件不变下的影响。这正是多元回归分析如此有用的原因所在。类似地，在保持 x_1 不变时

$$\Delta\hat{y}=\hat{\beta}_2\Delta x_2$$

例 3.1　大学 *GPA* 的决定因素

GPA1.RAW 中的变量包括了大学平均成绩（*colGPA*）、高中的平均成绩（*hsGPA*）和大学能力测验分数（*ACT*），样本是从一所规模较大的大学选取的 141 名学生；大学和高中的 *GPA* 都采取四分制。我们得到如下 OLS 估计线，用高中 *GPA* 和能力测验的分数来预测大学的 *GPA*：

$$\widehat{colGPA}=1.29+0.453hsGPA+0.0094\ ACT \tag{3.15}$$

我们该怎么解释这个方程呢？首先，1.29 是在 *hsGPA* 和 *ACT* 都为零时预测的大学 *GPA*。由于没有人能在高中 *GPA* 为零或能力测验得零分的情况下进入大学，所以在这个方程中，截距项本身并没有什么意义。

更有意义的估计值是 *colGPA* 和 *hsGPA* 的斜率系数。恰如所料，*colGPA* 和 *hsGPA* 之间存在正的局部相关：保持 *ACT* 不变，如果 *hsGPA* 提高 1 分，则大学 *GPA* 会提高 0.453 分或接近半分。换句话说，如果我们选择 A 和 B 两个学生，其 *ACT* 分数相同，但学生 A 的高中 *GPA* 比学生 *B* 的高中 *GPA* 高出 1 分，那么，我们预计学生 A 的大学 *GPA* 将比学生 *B* 的大学 *GPA* 高 0.453 分。（它没有提到任何两个真实学生，但却是我们最好的预测。）

74 *ACT* 的符号表明，在保持 *hsGPA* 不变时，*ACT* 分数变化 10 分（由于样本中 *ACT* 的平均分数也只有 24 分，且标准离差小于 3，所以，10 分的差距已经很大），对 *colGPA* 的影响还达不到1/10分。这个影响很小，而且表明，一旦采用高中 *GPA* 来解释大学 *GPA*，那么 *ACT* 分数就不能有力地预测大学 *GPA*。（很自然，虽然还有许多其他因素影响 *GPA*，但这里我们主要关注在高中生中容易得到的统计量。）在以后我们讨论了统计推断之后，我们将证明，*ACT* 的系数不仅实际上很小，而且还是统计上非显著的。

如果只将 colGPA 与 ACT 进行简单回归分析，则得到

$$\widehat{colGPA}=2.40+0.0271\,ACT$$

因此，ACT 的系数几乎是式（3.15）中估计值的 3 倍大。但这个方程使我们无法将两个具有同等高中 GPA 的同学进行比较，它对应于一个不同的实验。以后，我们还会更多地探讨多元和简单回归之间的差别。

不止两个自变量的情况与此相似。OLS 回归线

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2+\cdots+\hat{\beta}_kx_k \tag{3.16}$$

用变化量表示为

$$\Delta\hat{y}=\hat{\beta}_1\Delta x_1+\hat{\beta}_2\Delta x_2+\cdots+\hat{\beta}_k\Delta x_k \tag{3.17}$$

x_1 的系数度量的是，在所有其他条件不变的情况下，因提高一个单位的 x_1 而导致的变化。即在保持 x_2，x_3，…，x_k 不变的情况下，有

$$\Delta\hat{y}=\hat{\beta}_1\Delta x_1 \tag{3.18}$$

因此，在估计 x_1 对 y 的影响时，已经控制了变量 x_2，x_3，…，x_k 的影响。其他系数与此相似。

下面是一个含有三个自变量的例子。

例 3.2　小时工资方程

利用 WAGE1.RAW 对工人的 526 个观测数据，我们在解释 log（wage）的方程中包括了 educ（受教育年限）、exper（在劳动市场上的工作经历）和 tenure（现任职务的任期）。估计出来的方程是

$$\widehat{\log(wage)}=0.284+0.092educ+0.0041exper+0.022tenure \tag{3.19}$$

如在简单回归情形中一样，系数可做百分比解释。这里惟一的区别是，它们也可解释为在其他条件不变下的影响。系数 0.92 意味着在保持 exper 和 tenure 固定不变的情况下，多受一年教育者的 log（wage）预计会提高 0.092，即 9.2%。换句话说，如果有两个人具有同样的工作经历和任期，在教育水平相差一年时，educ 的系数表示了预期工资差别的百分比。对教育回报的这种度量至少保持两个重要的生产要素不变，无论它是否很好地估
75 计了多受一年教育在其他条件不变情况下的影响，我们都要研究一下 OLS 的统计特性（参见 3.3 节）。

多元回归中“保持其他因素不变”的含义

因为对多元回归分析中斜率参数的局部效应解释可能会导致一些混淆，

所以我们现在要尽量避免这个问题。

在例 3.1 中，我们观察到，*ACT* 的系数所度量的是在保持不变*hsGPA* 的情况下，预期 *colGPA* 的差别。多元回归分析的功能在于，尽管不能在其他条件不变的情况下搜集数据，但它提供的系数仍可做其他条件不变的解释。在对 *ACT* 的系数做局部效应解释时，看起来就好像我们出去在具有同等高中 *GPA* 但 *ACT* 分数可能不同的人群中抽样。然而情况并非如此，数据是来自一所很大的大学的随机样本：在获得数据的过程中，对 *hsGPA* 和 *ACT* 的样本值都没有施加任何限制。在获取样本时，我们很少奢侈到能限制某些变量不变的程度。如果我们能搜集到具有同等高中 *GPA* 的个人样本，那我们就能进行一个 *colGPA* 对 *ACT* 的简单回归分析。多元回归使我们有效地模拟了对自变量的值不加限制的情况。

多元回归分析使我们能在非实验环境中去做自然科学家在受控实验中所能做的事情：保持其他因素不变。

同时改变不止一个自变量

有时我们想改变一个以上的自变量，看看由此对因变量的影响。通过方程（3.17）很容易做到。例如，在方程（3.19）中，我们可以得到，当一个人在同一企业多待一年，即一般工作经历 *exper* 和 *tenure* 都增加一年时，对工资的估计影响。（保持 *educ* 不变下）总影响是

$$\Delta \widehat{\log(wage)} = 0.0041\Delta exper + 0.22\Delta tenure$$
$$= 0.0041 + 0.022 = 0.0261$$

或 2.6%。由于 *exper* 和 *tenure* 都增加一年，所以只要把 *exper* 和 *tenure* 的系数相加并乘以 100，就得到了总影响的百分数。

OLS 的拟合值和残差

在得到 OLS 回归线（3.11）后，我们对每次观测都得到一个**拟合值**或**预测值**。对观测 i，其拟合值就是

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} \qquad (3.20)$$

问题 3.2 ☞

76

在例 3.1 中，用高中 *GPA* 和 *ACT* 分数来解释大学 *GPA* 的 OLS 拟合线为：

$$\widehat{colGPA} = 1.29 + 0.453hsGPA + 0.0094ACT$$

如果平均的高中 *GPA* 约为 3.4，而平均的 *ACT* 分数约为 24.2，那么样本中平均的大学 *GPA* 是多少呢？

它只是通过将第 i 个自变量的值代入方程（3.11）所得到的预测值。在得到拟合值时，我们不应该忘记截距项；否则，结果就极具误导性。举例而言，如果在式（3.15）中，$hsGPA_i = 3.5$ 和 $ACT_i = 24$，那么 $i = 1.29 + 0.453(3.5) + 0.0094(24) = 3.101$（近似到小数点后三位）。

规范地讲，对任一观测 i，实际值都不等于预测值 $\hat{y}_i$；OLS 最小化了预测误差平方的平均值，但对任何一个观测的预测误差都没做说明。第 i 个观测的残差只是像在简单回归中那样，被定义为

$$\hat{u}_i = y_i - \hat{y}_i \tag{3.21}$$

每次观测都有一个残差。如果 $\hat{u}_i > 0$，则 $\hat{y}_i$ 小于 y_i，这意味着，对这个预测来说，y_i 被预测得过低；如果 $\hat{u}_i < 0$，则 $\hat{y}_i$ 大于 y_i，这意味着，对这个预测来说，y_i 被预测得过高。

OLS 拟合值和残差具有某些能直接从单变量情形推广而得到的重要性质：

1. 残差的样本平均值为零。

2. 每个自变量和 OLS 协残差之间的样本协方差为零。于是，OLS 拟合值和 OLS 之间的样本协方差也为零。

3. 点($\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k, \bar{y}$)总位于 OLS 回归线上：$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \cdots + \hat{\beta}_k \bar{x}_k$。

前面两个性质是用来得到 OLS 估计值的方程组的直接结果。方程组（3.13）中的第一个方程表明，残差和等于零。剩下两个方程的形式是 $\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0$，意味着每个自变量与 $\hat{u}_i$ 之间的协方差都是零。性质 3 从性质 1 可直接得到。

对多元回归"排除其他变量影响"的解释

在应用 OLS 时，我们不需要知道方程组（3.13）的解 $\hat{\beta}_j$ 的明确表达式。但为了进行某些推导，我们确实又需要知道 $\hat{\beta}_j$ 的明确表达式。这些表达式还能折射出 OLS 的工作方式。

再次考虑 $k = 2$ 个自变量的情形，$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$。为明确起见，我们只考虑 $\hat{\beta}_1$。一种表达 $\hat{\beta}_1$ 的方式为

77

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2} \tag{3.22}$$

式中，$\hat{r}_{i1}$ 为利用现有样本将 x_1 对 x_2 回归而得到的 OLS 残差。我们将第一个自变量 x_1 对第二个自变量 x_2 进行回归，然后得到残差（这里没有 y 的作用）。方程（3.22）表明，我们再将 y 对 $\hat{r}_1$ 进行简单回归就能得到 $\hat{\beta}_1$。（注

意到残差 $\hat{r}_{i1}$ 的样本均值为零，所以 $\hat{\beta}_1$ 就是通常从简单回归中得到的斜率参数。)

方程（3.22）中的表达式还给出 $\hat{\beta}_1$ 的另一个解释，即局部效应。残差 $\hat{r}_{i1}$ 是 x_{i1} 中与 x_{i2} 不相关的部分。另一种说法是，$\hat{r}_{i1}$ 是 x_{i1} 排除了 x_{i2} 的影响之后的部分。于是，$\hat{\beta}_1$ 度量了在排除 x_{i2} 的影响之后 y 和 x_1 之间的样本关系。

在简单回归分析中，由于回归中根本就没有其他变量，所以就不用排除其他变量的影响。习题 3.22 使用例 3.2 中的工资数据告诉你排除其他变量的过程。事实上重要的是，方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$ 中的 $\hat{\beta}_1$ 度量的是在保持 x_2 不变时，x_1 提高一个单位导致的 y 的变化量。

在一个含有 k 个解释变量的一般模型中，$\hat{\beta}_1$ 仍可写成方程（3.22），但残差 $\hat{r}_{i1}$ 来自 x_1 对 x_2，x_3，…，x_k 的回归。于是 $\hat{\beta}_1$ 度量的是在排除了 x_2，x_3，…，x_k 等变量的影响之后，x_1 对 y 的影响。

简单回归和多元回归估计值的比较

在两个特殊情形中，y 对 x_1 的简单回归所得到的回归估计值，等于将 y 对 x_1 和 x_2 做 OLS 回归时所得到 x_1 的偏回归估计值。更准确地说，将 y 对 x_1 的简单回归写作 $\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1x_1$，并将多元回归写作 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$。我们知道，简单回归系数 $\tilde{\beta}_1$ 通常并不等于多元回归系数 $\hat{\beta}_1$，但有两种明显的情况会使它们相等：

1. 样本中 x_2 对 y 的局部效应为零，即 $\hat{\beta}_2=0$。
2. 样本中 x_1 和 x_2 不相关。

第一个论断的证明，可以从决定 $\tilde{\beta}_0$，$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的两个方程看出：$\sum_{i=1}^{n}x_{i1}(y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1}-\hat{\beta}_2x_{i2})=0$ 和 $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}_1-\hat{\beta}_2\bar{x}_2$。令 $\hat{\beta}_2=0$，所得到的截距和斜率就是 y 对 x_1 的简单回归所得到的截距和斜率。

第二个论断来自方程(3.22)。如果样本中的 x_1 和 x_2 不相关，那么将 x_1 对 x_2 进行回归的结果发现，没有必要排除第二个变量的影响，所以 y 对 x_1 的简单回归与 y 对 x_1 和 x_2 的多元回归所得到关于 x_1 的偏回归估计值相同。

78 尽管简单回归和多元回归的估计值几乎从来都不会相等，但我们还是可以用以上的特性去解释它们为什么要么相去甚远，要么十分相似。例如，如果 $\hat{\beta}_2$ 很小，我们可能会预期 β_1 的简单回归和多元回归估计值很相似。在例 3.1 中，*hsGPA* 和 *ACT* 之间的样本相关系数是 0.346，这并不是一个微不足道的相关系数。但 *ACT* 的系数却相当小。*hsGPA* 对 *colGPA* 的简单回归得到的系数估计值为 0.482，与式（3.15）中的估计值 0.453 相差不远。

例 3.3　401k 养老金方案的参与

我们利用 401K.RAW 中的数据，估计了一种方案的贡献率 *mrate*（match rate）对在 401k 养老金方案中的参与率 *prate*（participation rate）的影

响。贡献率是指企业对一个工人所贡献的每1美元而向工人的养老基金贡献的数量（直到某个极限）；因此，$mrate = 0.75$ 意味着企业对工人贡献的每1美元都向基金贡献75美分。参与率则是有资格拥有一个401k账户的工人中参与此方案的百分比。变量 age 是401k养老金方案的年岁。数据集中有1 534个方案，平均的 $prate$ 为87.35，平均的 $mrate$ 为0.732，而平均的 age 为13.2。

将 $prate$ 对 $mrate$ 和 age 回归，得到

$$\widehat{prate} = 80.12 + 5.52 mrate + 0.243 age \tag{3.23}$$

可见，$mrate$ 和 age 都具有预期的效应。如果我们不控制 age 会怎么样呢？由于 age 的估计影响并非微不足道，所以预期在从回归方程中去掉 age 后，$mrate$ 的估计影响会有很大变动。但将 $prate$ 对 $mrate$ 进行简单回归得到 $\widetilde{prate} = 83.08 + 5.86 mrate$。虽然 $mrate$ 对 $prate$ 之影响的简单回归估计值与多元回归估计值明显不同，但相差不大。（简单回归估计值只比多元回归估计值大6.2%。）这种情况，可由 $mrate$ 与 age 之间的样本相关程度只有0.12这个事实来解释。

在含有 k 个自变量的情形中，只有在如下两种情况下，y 对 x_1 进行简单回归与 y 对 x_1，x_2，…，x_k 进行多元回归所得到关于 x_1 的估计值才会相同：(1)从 x_2 到 x_k，所有的OLS系数都是零，或者(2) x_1 与 x_1，x_2，…，x_k 都不相关。实际上，这两个条件都不太可能成立。但如果所有从 x_2 到 x_k 的系数都很小，或者 x_1 与其他自变量之间的样本相关关系不显著，那么 x_1 对 y 的影响的简单回归估计值和多元回归估计值可能会很相似。

拟合优度

像简单回归一样，我们可以定义**总平方和**（SST）、**解释平方和**（SSE）和**残差平方和**（SSR）分别为

$$\mathrm{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{3.24}$$

79

$$\mathrm{SSE} \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{3.25}$$

$$\mathrm{SST} \equiv \sum_{i=1}^{n}\hat{u}_i^2 \tag{3.26}$$

利用简单回归情形中同样的论证，我们可以证明

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR} \tag{3.27}$$

换句话说，$\{y_i\}$ 中的总变异是 $\{\hat{y}_i\}$ 和 $\{\hat{u}_i\}$ 的总变异之和。

假定 y 的总变异不为零（除非在样本中的 y_i 是常数的情况下），我们可

以将式（3.27）的两边同时除以 SST，得到

$$\frac{\mathrm{SSE}}{\mathrm{SST}}+\frac{\mathrm{SSR}}{\mathrm{SST}}=1$$

恰如在简单回归中一样，被定义为

$$R^2 \equiv \frac{\mathrm{SSE}}{\mathrm{SST}}=1-\frac{\mathrm{SSR}}{\mathrm{SST}} \tag{3.28}$$

而且被解释为样本方差中被 OLS 回归线所解释的部分。根据定义，R^2 是一个介于 0～1 之间的数。

还可以证明，R^2 等于 y_i 的实际值及其拟合值 $\hat{y}_i$ 之间的相关系数的平方。即

$$R^2=\frac{\left[\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\right]^2}{\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right]\left[\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2\right]} \tag{3.29}$$

[在式(3.29)中，我们已将 $\hat{y}_i$ 的平均值表示成相关系数所需要的形式；但由于残差的样本平均值为零，而且 $y_i=\hat{y}_i+\hat{u}_i$，所以这个平均值 $\bar{\hat{y}}_2$ 就等于 $\bar{y}$。]

有关 R^2 的一个重要事实是，在回归中多增加一个自变量后，它绝对不会减小，而且通常会增大。这个代数命题之所以成立，是因为在模型中多增加一个回归元时，按照定义，残差平方和绝对不会增加。

回归中增加任何一个变量都不会使 R^2 减小的事实，使得用 R^2 作为决定在模型中增加一个或几个变量的手段很不适当。决定一个解释变量是否应放入模型的因素是，这个解释变量在**总体**中对 y 的局部效应是否非零。我们将在第 4 章讨论统计推断时说明如何对这个假设进行检验。我们还将看到，如果应用适当，我们还能使用 R^2 来检验一组变量，以看其在解释 y 中是否重要。就目前而言，我们还只把它作为对给定模型拟合优度的一种度量。

例 3.4　大学 *GPA* 的决定因素

80 *从我们前面所考虑的大学平均成绩模型得到的回归方程为*

$$\widehat{colGPA}=1.29+0.453\,hsGPA+0.0094\,ACT$$
$$n=141,\ R^2=0.176$$

这意味着，hsGPA 和 ACT 一起解释了这个学生样本中大学 GPA 方差的 17.6%。虽然看上去这个比例不是很高，但我们必须记得，能影响一个学生大学表现的因素还有很多，包括家庭背景、个性、高中教育的质量和对大学的喜恶等。如果 hsGPA 和 ACT 解释了 GPA 方差的几乎全部，那么在大学的表现将由高中的表现预先决定了。

例 3.5　对拘捕记录的解释

CRIME1.RAW 包含了加利福尼亚 1960 年或 1961 年出生的 2 725 名男

子在 1986 年的拘捕数据和其他信息。在 1986 年以前，样本中的每一个人都至少被拘捕过一次。变量 *narr*86 表示这个人在 1986 年被被拘捕的次数，对样本中的多数人（72.29%）来说，它等于零，而且变化范围是 0～12。（在 1986 年被捕的比例是 20.51%。）变量 *pcnv* 表示 1986 年前被捕导致定罪的比例（不是百分数），*avgsen* 表示此前定罪被宣判的平均时间长度（多数人都是零），*ptime*86 表示 1986 年在监狱里度过的月数，而 *qemp*86 则表示此人在 1986 年被雇用的季度数（从 1～4）。

一个解释拘捕的线性模型为

$$narr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 ptime86 + \beta_4 qemp86 + u$$

式中，*pcnv* 为犯罪被定罪的可能性；*avgsen* 度量了如果被定罪，预期惩罚的严重性；变量 *ptime*86 刻画的是犯罪的监禁效应：如果一个人呆在监狱中，那他就不可能因监外的一桩犯罪而被捕；劳动市场的机会由 *qemp*86 粗略刻画。

首先，在没有变量的情况下估计这个模型，我们得到

$$\widehat{narr86} = 0.712 - 0.15 pcnv - 0.34 ptime86 - 0.104 qemp86$$

$$n = 2\ 725, R^2 = 0.041\ 3$$

这个方程表明，作为一个整体，*pcnv*、*ptime*86 和 *qemp*86 三个变量解释了 *narr*86 的方差的 4.1%。

每个 OLS 斜率系数都具有预料中的符号。定罪比率的提高会降低预期的拘捕次数。如果我们将 *pcnv* 提高 0.50（定罪概率的很大提高），那么，保持其他因素不变，$\Delta\widehat{narr86} = -0.150(0.5) = -0.075$。由于拘捕不能只变化一个分数，所以这看起来不太习惯，但我们可以用这个数值得到一大群人预期拘捕的变化。例如，在 100 个人中，当 *pcnv* 提高 0.50 时，预计拘捕会减少 7.5 次。

81 类似地，监禁期越长，预计的拘捕次数越低。事实上，如果 *ptime*86 从 0 提高到 12，对一个特定的人来说，预计的拘捕次数会减少 $0.034(12) = 0.408$。延长一个季度的合法就业，会使预计的拘捕次数减少 0.104，对 100 个人来说，将减少 10.4 次。

如果模型中增加 *avgsen*，我们知道 R^2 将增加。被估计的方程是

$$\begin{aligned}\widehat{narr86} = {} & 0.707 - 0.151 pcnv + 0.007\ 4 avgsen \\ & - 0.037 ptime86 - 0.103 qemp86 \\ & n = 2\ 725, R^2 = 0.042\ 2\end{aligned}$$

于是，增加变量 *avgsen* 使 R^2 从 0.041 3 增加到 0.422，这是一个特别小的影响。*avgsen* 系数的符号也是出乎意料的：判刑时间较长，则会增加犯罪活动。

对例 3.5 最后还有一句话值得告诫。第二个回归中包含 4 个解释变量也只解释了 *narr*86 的方差的 4.2%，这个事实并不意味着这个方程没有用

处。尽管所有这些解释变量解释不了大部分的拘捕方差，但OLS估计值仍有可能成为每个解释变量在其他条件不变情况下对 *narr*86 之影响的可靠估计值。如我们将会看到的那样，是不是可靠并不直接取决于 R^2 的大小。一般来说，较低的 R^2 表明，很难准确地预计 y 的个别结果。我们将在第6章更详尽地研究这个问题。在拘捕一例中，较小的 R^2 反映了我们已在社会科学中所预想的东西：它通常很难预计个别人的行为。

通过原点的回归

有时，一种经济理论或常识会告诉我们，β_0 应该为零，所以应该简要地提一下在截距为零时的OLS估计。具体而言，我们现在寻求一个具有如下形式的方程：

$$\tilde{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \cdots + \tilde{\beta}_k x_k \tag{3.30}$$

式中，估计值上面的符号“～”，以区别于带截距的OLS回归［如式(3.11)］。在式(3.30)中，当 $x_1 = 0, x_2 = 0, \cdots, x_k = 0$ 时，则预测值也为零。在这种情况下，$\tilde{\beta}_1, \cdots, \tilde{\beta}_k$ 被称为从 y 对 $x_1, x_2, \cdots, x_k$ 进行**通过原点**的回归而得到的OLS估计值。

像往常一样，式(3.30)中的OLS估计值最小化了残差平方和，只是截距项被设定为零。应该注意，我们以前推导的OLS的性质，对通过原点的回归不再成立。特别是，OLS残差的样本平均不再为零。而且，如果 R^2 被定义为 $R^2 = 1 - \mathrm{SSR/SST}$，其中SST由式(3.24)给出，而SSR现在是 $\sum_{i=1}^{n}(y_i - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_k x_{ik})^2$，于是，$R^2$ 实际上可能为负，这意味着样本平均比解释
82 变量更多地“解释”了 y_i 的方差。我们要么应该在回归中包含一个截距项，要么断定解释变量对 y 的解释很差。为了使 R^2 总是非负，有些经济学家更喜欢像在式(3.29)中一样，用 y 的实际值和拟合值之间相关系数的平方来计算 R^2。（在此情形下，由于拟合值的平均值不再等于 $\bar{y}$，所以必须直接计算。）但对于通过原点的回归，计算 R^2 没有一套特定的规则。

通过原点的回归有一个重要的缺陷，如果总体模型中的截距项 β_0 不为零，那么斜率参数的OLS估计量将是有偏的。在某些情况下，这种偏误可能会很严重。当 β_0 确实为零时，估计带截距项方程的代价是OLS斜率估计量的方差会更大。

3.3 OLS 估计量的期望值

现在转而讨论在估计一个产生样本的总体模型的参数时OLS所具有的

统计性质。我们推导 OLS 估计量的期望值。特别是，说明并讨论四个假定，这些假定都是对简单回归模型假定的直接推广，而且在这些假定之下，OLS 估计量是总体参数的无偏估计。我们还能明确得到在回归中漏掉一个重要变量时的偏误。

应该记得，当反复进行随机抽样时，统计性质与一个特定的样本无关，而与估计量的性质有关。因此，3.3 节、3.4 节和 3.5 节多少有点抽象。当我们举例推导特定模型的偏误时，谈论从单一个样本中得到的一系列估计值的统计特性是没有意义的。

下面所做的第一个假定不外是给多元线性回归（MLR）模型下一个定义。

假定 MLR.1（对参数而言为线性）

总体模型可写成

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k+u \qquad (3.31)$$

式中，β_0，β_1，⋯，β_k 为我们所关心的未知参数（常数）；u 为观测不到的随机误差或随机干扰项。

方程（3.31）规范地表述了**总体模型**或**真实模型**（population model or true model），它使我们有可能估计一个与方程（3.31）不同的模型。此模型的一个重要特点是，它是参数 β_0，β_1，⋯，β_k 的线性函数。如我们所知，由于 y 和自变量都可以是我们所关心的变量的任意函数（如自然对数函数和平方等），所以方程（3.31）相当灵活多变［例如参见方程（3.7）］。

假定 MLR.2（随机抽样性）

我们有一个含 n 次观测的随机样本 $\{(x_{i1},\ x_{i2},\ \cdots,\ x_{ik},\ y_i):i=1,2,\ \cdots,\ n\}$，它来自由方程（3.31）描述的总体模型。

83 有时我们需要对一次特定的观测（从总体中随机抽取的一次观测 i）写出它的方程

$$y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_kx_{ik}+u_i \qquad (3.32)$$

记住 i 表示观测次第，x 的第二个角标表示变量序号。例如，我们可以对某个特定的 CEO 写出其 CEO 薪水方程为

$$\log(salary_i)=\beta_0+\beta_1\log(sales_i)+\beta_2ceoten_i+\beta_3ceoten_i^2+u_i \qquad (3.33)$$

式中，u_i 包含了影响第 i 个 CEO 的薪水而又观测不到的因素。就应用而言，通常最容易的是像方程（3.31）那样写出总体模型。它不是那么难懂，从而只强调了我们所关心的是估计总体关系。

借助模型（3.31），从 y 对 $x_1,x_2,\cdots,x_k$ 的回归中得到的 OLS 估计量 $\hat{\beta}_0$，$\hat{\beta}_1,\cdots,\hat{\beta}_k$，现在被看做 $\beta_0,\beta_1,\cdots,\beta_k$ 的估计量。我们在 3.2 节看到，OLS 对一个特定样本选择的估计值，会使残差的平均为零，每个自变量和残差的样本相

关也为零。要使 OLS 无偏，需要这个条件对总体来说也成立。

假定 MLR.3（条件均值为零）

给定自变量的任何值，误差 u 的期望值为零。换句话说，即

$$E(u \mid x_1, x_2, \cdots, x_k) = 0 \tag{3.34}$$

假定 MLR.3 可能不成立的情况之一是，方程（3.31）中被解释变量和解释变量之间的函数关系被错误地设定。例如，如果我们在估计消费模型时，在消费函数 $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$ 中忘掉了二次项 inc^2。发生另外一种函数形式误设的情况是，当一个变量在总体中应该以对数出现时我们却使用了其水平值，或者相反。比如，如果真正的模型是以 log（*wage*）作为因变量，但在回归分析中却以 *wage* 作为因变量，那么估计量就是有偏的。从直觉上看，这应该相当明显。在第 9 章我们将讨论侦破函数形式误设的方法。

漏掉一个与 x_1，x_2，…，x_k 中任何一个相关的重要因素也能导致假定 MLR.3 不成立。使用多元回归分析，我们能包含解释变量中的许多因素，漏掉一些变量在多元回归分析中发生的可能性比在简单回归分析中要小很多。尽管如此，由于数据局限或被忽略，在任何一个应用研究中，总有一些因素我们不能包括进来。如果我们认为这些因素应该加以控制，而且它们与一个或多个自变量相关，那么这样一来就会违背假定 MLR.3。我们稍后在一些简单模型中推导这个偏误。

84 u 还可能有其他方式与一个解释变量相关。在第 15 章，我们将讨论解释变量的测量误差的问题。在第 16 章，我们将讨论一些在概念上更困难的问题，其中一个或多个解释变量与 y 同时决定。在我们对一系列理想的假定条件之下的多元回归分析有清楚的了解之前，必须推迟对这些问题的研究。

当假定 MLR.3 成立时，我们通常称我们具有**外生解释变量**（exogenous explanatory variables）。如果出于某种原因 x_j 仍与 u 相关，那么 x_j 就被称为**内生解释变量**（endogenous explanatory variable）。虽然“外生”和“内生”的术语源自联立方程分析（参见第 16 章），但“内生解释变量”一词涵盖了一个解释变量可能与误差项相关的一切情况。

证明 OLS 的无偏性所需要的最后一个假定是，保证 OLS 估计量确实有完好的定义。对于简单回归，我们需要假定样本中唯一的一个自变量不是一个常数。在多元回归分析中对应的假定则要复杂得多。

假定 MLR.4（不存在完全共线性）

在样本（因而在总体中），没有一个自变量是常数，自变量之间也不存在严格的线性关系。

不存在完全共线性的假定只考虑了自变量。由于初学计量经济学的学生总是将假定 MLR.3 与假定 MLR.4 相混淆，所以我们在此强调，MLR.4 对 u 与解释变量之间的关系只字未提。

由于我们现在必须关注所有自变量之间的关系，所以假定 MLR.4 比在

简单回归情况下更复杂。如果方程（3.31）中的一个自变量刚好是其他自变量的一个线性组合，那么就说这个模型遇到**完全共线性**（perfect collinearity）的问题，也就不能由OLS来估计。

重要的是要注意到，假定MLR.4**允许**自变量之间存在相关关系，只是不能完全相关。如果我们不容许自变量之间存在任何相关关系，那么多元回归分析对计量经济分析就没有多大用处。例如，在考试分数与教育支出和平均家庭收入相关的模型中

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$$

我们充分预料到 $expend$ 和 $avginc$ 之间可能相关：学生家庭平均收入高的学校，倾向于对每个学生在教育上支出更多。实际上，在方程中包括变量 $avginc$ 的主要动机是，我们怀疑它与 $expend$ 之间存在相关关系，所以我们可能在分析中将它保持不变。假定MLR.4只是排除了在我们的样本中 $expend$ 和 $avginc$ 之间完全相关的情况。如果得到一个学生平均支出与平均家庭收入完全相关的样本，那我们就太不幸了。但可以预见到，这两个自变量之间存在某些相关，还可能是很大的相关，而且这当然是允许的。

85 两个自变量完全相关最简单的方式是，一个变量是另一个变量的常数倍。当研究者不小心将同一个变量在不同的度量单位下两次进入同一个回归方程时，就有可能发生这种情况。例如，在估计消费与收入的关系时，将收入以美元为单位和以千美元为单位分别作为自变量就毫无意义，其中有一个自变量是多余的。在保持以美元为单位度量的收入不变时，改变以千美元为单位的收入又有什么意义呢？

我们已经知道，同一变量不同的非线性函数也可以出现在回归元中。比方说，模型 $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$ 就不违背假定MLR.4：尽管 $x_2 = inc^2$ 刚好是 $x_1 = inc$ 的一个函数，但 inc^2 并非刚好是 inc 的一个线性函数。与在模型中包括以美元度量的收入和以千美元度量的收入不同，在模型中包括 inc^2 是推广函数形式的一种有用方法。

常识告诉我们，不要在同一个回归方程中包括以不同单位度量的同一个解释变量。一个自变量也可能以更微妙的方式成为另一个自变量的倍数。假设我们想估计一个常弹性消费函数的扩展形式。看来设定一个模型

$$\log(cons) = \beta_0 + \beta_1\log(inc) + \beta_2\log(inc^2) + u \qquad (3.35)$$

很自然，其中 $x_1 = \log(inc)$，而 $x_2 = \log(inc^2)$。利用自然对数的基本性质（参见附录A），我们得到 $\log(inc^2) = 2\log(inc)$，即 $x_2 = 2x_1$。而且自然而然地，这对样本中的每一个观测都成立。这就违背了假定MLR.4。我们应该做的是，用 $[\log(inc)]^2$ 来替换 $\log(inc^2)$。这是对常弹性模型的一种合理扩展，我们将在第6章看出如何解释这种模型。

自变量可能完全线性相关的另一种方式是，一个自变量恰好可以表达成其他两个或多个自变量的一个线性函数。例如，假设我们想估计竞选支出对竞选结果的影响。为简便起见，假定每次选举都有两位候选人。令 $voteA$ 为

候选人 A 的得票率，*expendA* 为候选人 A 的竞选支出；*expendB* 为候选人 B 的竞选支出，并令 *totexpend* 为竞选总支出；后面三个变量都以美元为单位度量。为了将每个候选人竞选支出与总支出的影响隔离开来，将模型做如下设定看来也很自然：

$$voteA = \beta_0 + \beta_1 expendA + \beta_2 expendB + \beta_3 totexpend + u \qquad (3.36)$$

但由于根据定义，$x_3 = x_1 + x_2$，所以这个模型违背了假定 MLR.4。试图在其他条件不变模式下解释这个方程就会揭示出这个问题。方程（3.36）中的参数 β_1 被认为是在保持候选人 B 的竞选支出和总支出不变的情况下，度量了候选人 A 竞选支出的增加对其得票率的影响。因为如果 *expendB* 和 *totexpend* 都保持不变，我们就不可能增加 *expendA*，所以这就毫无意义。

86 对于方程（3.36）中的完全共线性，解决办法很简单：从模型中去掉三个变量中的任意一个。我们可能会去掉 *totexpend*，那么 *expendA* 的系数就度量着在保持 B 的竞选支出不变的情况下，A 增加竞选支出对其得票率的影响。

问题 3.3

在前面这个例子中，如果我们使用 *expendA*，*expendB* 和 *shareA* 作为解释变量，其中 $shareA = 100(expendA / totexpend)$ 是候选人 A 的竞选支出占竞选总支出的比例，这样会违背假定 MLR.4 吗？

前面的例子表明，如果我们在设定模型时不小心从事，假定 MLR.4 就有可能不成立。在样本容量 n 相对于被估计的参数个数而言太小的话，假定 MLR.4 也会不成立。在方程（3.31）这个一般回归模型中，如果样本容量小于参数个数 $k+1$，MLR.4 就不成立。从直觉上讲，这合乎情理：要估计 $k+1$ 个参数，我们至少需要 $k+1$ 个观测。无足为奇，我们在 3.4 节计算方差时将会看到，最好是能得到尽可能多的观测值。

如果正确地设定了模型，而且 $n \geqslant k+1$，在因为搜集样本时运气不佳而导致的极少情况下，假定 MLR.4 也会不成立。比如，在一个以受教育程度和工作经历为变量的工资方程中，有可能在我们得到的一个随机样本中，每个人的受教育年限都正好是其工作经历年限的 2 倍。虽然这种情况将导致假定 MLR.4 不能成立，但可以认为，除非样本容量极小，否则很不可能。

我们现在准备证明，在上述四个多元回归假定之下，OLS 估计量是无偏的。如在简单回归情形中一样，这些估计量的期望值要视样本中自变量的值而定，但我们不显式地表明这种条件关系。

定理 3.1（OLS 的无偏性）

在假定 MLR.1～MLR.4 下，下式对总体参数的任意值都成立：

$$E(\hat{\beta}_j) = \beta_j, \quad j = 0, 1, \cdots, k \qquad (3.37)$$

即 OLS 估计量是总体参数的无偏估计量。

在我们前面实证的例子中，假定 MLR.4 已经满足（因为已经能够计算

OLS 估计值)。此外，多数情况下，样本都是从一个定义完好的总体中随机抽取的。如果我们相信所设定的这些模型在关键假定 MLR.3 下是正确的，就能断定，在这些例子中，OLS 是无偏的。

由于我们正逐渐到达在严肃的实证研究中能使用多元回归的地步，所以回忆一下无偏的含义是有帮助的。在像方程（3.19）中工资方程那样的例子中，它的含义就是说明“9.2%是教育回报的一个无偏估计值”等。如我们所知，一个估计值不可能是无偏的：一个估计值就是从一个特定的样本得到的一个固定的数，它通常不等于总体参数。我们说 OLS 在假定 MLR.1～ 87 MLR.4 下是无偏的，此时是指当将用来得到 OLS 估计值的**程序**用于各种可能的随机样本时，这个程序是无偏的。虽然希望我们已经得到一个样本，其给出的估计值能接近总体值，但不幸的是，这不可能得到保证。

在回归模型中包含了无关变量

我们能很快解决的一个问题是，在多元回归分析中**包含了一个无关变量**（inclusion of an irrelevant variable）或**对模型进行了过度设定**（overspecifying the model）。也就是说，尽管一个（或多个）自变量在总体中对 y 没有影响，却被放到了模型中。(即它的总体系数为零。)

为了说明这个问题，假设我们将模型设定为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \tag{3.38}$$

而且这个模型满足假定 MLR.1～MLR.4。但在控制了 x_1 和 x_2 之后，x_3 对 y 没有影响，即 $\beta_3 = 0$。变量 x_3 与 x_1 和 x_2 或许相关或许不相关，问题是，一旦控制了 x_1 和 x_2，x_3 对 y 就没有影响。用条件期望的术语来说就是 $\mathrm{E}(y \mid x_1, x_2, x_3) = \mathrm{E}(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$。

因为不知道 $\beta_3 = 0$，所以倾向于估计包括 x_3 的方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 \tag{3.39}$$

在回归中包含了无关变量 x_3。当 x_3 在总体模型（3.38）中的系数为零时，在式（3.39）中包含了 x_3 会有什么影响呢？就 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的无偏性而言，不存在什么影响。因为这个结论直接从定理 3.1 得到，所以无须特别推导。记住，无偏性意味着对 β_j 的任意值（包括 $\beta_j = 0$），都有 $\mathrm{E}(\hat{\beta}_j) = \beta_j$。于是，我们断定 $\mathrm{E}(\hat{\beta}_0) = \beta_0, \mathrm{E}(\hat{\beta}_1) = \beta_1, \mathrm{E}(\hat{\beta}_2) = \beta_2$，以及 $\mathrm{E}(\hat{\beta}_3) = 0$（对 β_0, β_1 和 β_2 的任何值都成立）。尽管 $\hat{\beta}_3$ 本身不可能恰好为零，但其对许多随机样本的平均值将是零。

前面这个例子的结论是相当一般性的：在一个多元回归模型中包含一个或多个无关变量，或对模型进行了过度设定，并不会影响到 OLS 估计量的无偏性。这是否意味着包含无关变量的做法就没有危害呢？不是，如我们在 3.4 节将看到的那样，包含无关变量对 OLS 估计量的方差具有不利的影响。

遗漏变量的偏误：简单情形

现在假设，不是包含了一个无关变量，而是遗漏了一个实际上应包括在真实（或总体）模型中的变量。这通常被称为**排除一个有关变量**（excluding a relevant variable）或**对模型设定不足**（underspecifying the model）的问题。我们在第 2 章和本章的前半部分指出，这个问题一般会导致 OLS 估计量产生偏误。现在是明确证明这一点的时候了，而且同样重要的是推导出偏误的方向和大小。

88 推导遗漏一个重要变量导致的偏误是**误设分析**（misspecification analysis）的一个例子。我们从真实的总体模型具有两个解释变量和一个误差项的情况开始：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{3.40}$$

而且假定这个模型满足假定 MLR.1～MLR.4。

假定我们主要关心的是 x_1 对 y 的局部影响 β_1。比如，y 是小时工资（或小时工资的对数），x_1 是受教育水平，而 x_2 则是对天生能力的一种度量。为了得到 β_1 的一个无偏估计量，应该将 y 对 x_1 和 x_2 进行回归（这样就给出 β_0，β_1 和 β_2 的无偏估计量）。但由于我们的疏忽或数据不足，所以在排除 x_2 的情况下估计了这个模型。换句话说，我们只是将 y 对 x_1 进行了简单回归，得到方程

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \tag{3.41}$$

我们使用符号“～”而不是“^”是为了强调 $\tilde{\beta}_1$ 是来自于一个设定不足的模型。

初学遗漏变量的问题时，学生很难区别潜在的真实模型［这里是模型（3.40）］和我们实际估计的方程［这里由模型（3.41）中的回归来刻画］。虽然初看起来变量 x_2 应该放在模型中，而将它漏掉是很愚蠢的，但通常我们别无选择。比如，假设 *wage* 由

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + u \tag{3.42}$$

决定。由于能力不可观测，所以我们转而估计了模型

$$wage = \beta_0 + \beta_1 educ + v$$

式中，$v = \beta_2 abil + u$。将 *wage* 对 *educ* 进行简单回归，由此得到 β_1 的估计量就是我们所说的 $\tilde{\beta}_1$。

我们推导的 $\tilde{\beta}_1$ 的期望值要视 x_1 和 x_2 的样本值而定。因为 $\tilde{\beta}_1$ 恰好是简单回归中 OLS 斜率估计量，而且我们已在第 2 章深入地研究过它，所以推导出这个期望值并不困难。这里的不同之处在于，我们必须分析简单回归模型因遗漏一个变量而误设时所具有的性质。

通过方程（2.49），我们可以将 $\tilde{\beta}_1$ 表达成

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2} \tag{3.43}$$

下一步是最重要的一步。由于模型（3.40）是真实模型，所以对每个观测 i，我们将 y 写成

89
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \tag{3.44}$$

（因为真实模型包含了 x_2，所以不是 $y_i = \beta_0 + \beta_1 x_{i1} + u_i$）。令 SST_1 表示式（3.43）中的分母。如果我们将式（3.44）中的 y_i 代入式（3.43），则式（3.43）中的分子变成

$$\begin{aligned}&\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+u_i)\\&=\beta_1\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2+\beta_2\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)x_{i2}+\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)u_i\\&\equiv\beta_1\mathrm{SST}_1+\beta_2\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)x_{i2}+\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)u_i\end{aligned} \tag{3.45}$$

如果将式（3.45）的两边同时除以 SST_1，再根据自变量的值对上式求期望，并利用 E（u_i）＝0，就得到

$$\mathrm{E}(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2} \tag{3.46}$$

因此，$\mathrm{E}(\tilde{\beta}_1)$一般都不等于 β_1：$\tilde{\beta}_1$ 相对 β_1 而言是有偏的。

对于式(3.46)中 β_2 的系数，有一个简单的解释：如果使用我们的自变量样本，将 x_2 对 x_1 进行回归，它刚好是回归方程中的斜率系数，可以将这个回归方程写成

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1 \tag{3.47}$$

由于我们对两个自变量的样本值作出规定，所以这里的 $\tilde{\delta}_1$ 不是随机的。于是，可以将式（3.46）写成

$$\mathrm{E}(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1 \tag{3.48}$$

它意味着 $\tilde{\beta}_1$ 的偏误是$\mathrm{E}(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$，通常被称为**遗漏变量偏误**(omitted variable bias)。

我们从方程（3.48）可以看出，有两种情况使 $\tilde{\beta}_1$ 是无偏的。第一种情况相当显然：如果 $\beta_2 = 0$——即 x_2 不会出现在真实模型中，那么 $\tilde{\beta}_1$ 就是无偏的。我们已从第 2 章的简单回归分析中了解到了这一点。第二种情况更有意思，如果 $\tilde{\delta}_1 = 0$，即使 $\beta_2 \neq 0$，$\tilde{\beta}_1$ 相对 β_1 而言也是无偏估计。

由于 $\tilde{\delta}_1$ 是 x_1 和 x_2 之间的样本协方差与 x_1 的样本方差之比，所以当且仅当样本中的 x_1 和 x_2 不相关，才会有 $\tilde{\delta}_1 = 0$。于是，我们就有了一个

重要结论，如果样本 x_1 和 x_2 不相关，那 $\tilde{\beta}_1$ 就是无偏估计。这无足为奇，在 3.2 节我们就证明了，当样本中的 x_1 和 x_2 不相关时，简单回归估计量 $\tilde{\beta}_1$ 与多元回归估计量 $\hat{\beta}_1$ 相等。[我们还能证明，如果 $\mathrm{E}(x_2|x_1)=\mathrm{E}(x_2)$，那么 $\tilde{\beta}_1$ 的无偏性无须以 x_{i2} 为条件；于是在估计 β_1 时，只要调整截距，将 90 x_2 放在误差项中并不违背误差项的条件均值为零的假定。]

当 x_1 和 x_2 相关时，$\tilde{\delta}_1$ 与 x_1 和 x_2 之间的相关系数具有相同的符号：如果 x_1 和 x_2 正相关，则 $\tilde{\delta}_1>0$；而如果 x_1 和 x_2 负相关，则 $\tilde{\delta}_1<0$。$\tilde{\beta}_1$ 偏误的符号同时取决于 β_2 和 $\tilde{\delta}_1$ 的符号。表 3.2 总结了存在偏误时的四种可能情形。对表 3.2 需要仔细研究。比如，如果 $\beta_2>0$（x_2 对 y 有正影响），而 x_1 和 x_2 正相关，则 $\tilde{\beta}_1$ 的偏误就为正。如果 $\beta_2>0$，而 x_1 和 x_2 负相关，则偏误为负。如此等等。

表 3.2　在估计方程（3.40）的过程中遗漏了变量 x_2 时导致的可能偏误之总结

	$\mathrm{Corr}(x_1,x_2)>0$	$\mathrm{Corr}(x_1,x_2)<0$
$\beta_2>0$	偏误为正	偏误为负
$\beta_2<0$	偏误为负	偏误为正

虽然表 3.2 总结了偏误的方向，但偏误的大小仍然十分重要。偏误很小，无论符号是正是负，都不值得考虑。例如，如果总体中的教育回报是 8.6%，而 OLS 估计量中的偏误是 0.1%（1/10 个百分点），那我们就不是很关心。相反，偏误如果达到约 3 个百分点，问题就很严重了。偏误的大小由 β_2 和 $\tilde{\delta}_1$ 的大小决定。

实践中，由于 β_2 是一个未知的总体参数，所以就不能肯定 β_2 是正还是负。但就 x_2 对 y 产生影响的方向而言，我们通常都有一个不错的概念。此外，尽管 x_1 和 x_2 之间相关关系的符号因不能观测到 x_2 而未知，但多数情况下，我们对它们之间是正相关还是负相关总能作出有根据的猜测。

在工资方程（3.42）中，根据定义，更强的能力会导致更高的生产力，因而带来更高的工资：$\beta_2>0$。而且有理由相信 *educ* 和 *abil* 正相关：一般来说，越有天赋的人就选择越高的受教育水平。因此，从简单回归方程 $wage=\beta_0+\beta_1 educ+v$ 得到的 OLS 估计值一般都太大。这并不意味着从我们的样本所得到的估计值也太大。我们只能说，如果搜集许多随机样本，而且每次都得到简单回归估计值，那么这些估计值的平均值将比 β_1 更大。

例 3.6　小时工资方程

假设模型 $\log(wage)=\beta_0+\beta_1 educ+\beta_2 abil+u$ 满足假定 MLR.1－MLR.4。但 WAGE1.RAW 中的数据集没有包含能力方面的数据，所以我们 91 用如下简单回归来估计 β_1：

$$\log(\widetilde{wage}) = 0.584 + 0.083educ$$

$$n = 526,\ R^2 = 0.186$$

由于这只是从一个样本得到的结论，所以我们不能说 0.083 就比 β_1 更大；教育的真实回报可能低于也可能高于 8.3%（而且我们永远也不会确切地知道）。不过我们知道的是，对所有随机样本来说，估计值的平均将太大。

作为第二个例子，假设学生在初中水平标准化考试的平均分数由下式决定：

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 povrate + u$$

式中，*expend* 为每个学生的平均支出；*povrate* 为学校学生的贫穷率。使用学区数据时，只有学生及格比例和学生人均支出的观测值，没有贫穷率方面的信息。因此，用 *avgscore* 对 *expend* 的简单回归来估计 β_1。

我们又能得到 $\widetilde{\beta}_1$ 类似的偏误。首先，β_2 可能为负：大量的证据表明，生活在贫穷中的孩子总体上标准化考试的分数偏低。其次，每个学生的平均支出可能与贫穷率负相关：贫穷率越高，每个学生的平均支出就越低，所以 Corr（x_1、x_2）<0。从表 3.2 看，$\widetilde{\beta}_1$ 将产生正偏误。这一察觉具有重要的含义。支出的真实效应可能为零，即 $\beta_1 = 0$。但 β_1 的简单回归估计值通常都大于零，而且这会使我们在支出并不重要时却断定它们重要。

在阅读和实施经济学中的经验研究时，掌握与偏误估计量相联系的术语是很重要的。在模型（3.40）中漏掉一个变量的背景下，如果 $E(\widetilde{\beta}_1) > \beta_1$，那么我们就说 $\widetilde{\beta}_1$ **有向上的偏误**（upward bias）。当 $E(\widetilde{\beta}_1) < \beta_1$ 时，则 $\widetilde{\beta}_1$ **有向下的偏误**（downward bias）。这些定义与 β_1 的值为正或负如出一辙。**向零的偏误**（biased toward zero）一语是指 E（$\widetilde{\beta}_1$）比 β_1 更接近于零的情况。因此，如果 β_1 为正，那么 $\widetilde{\beta}_1$ 向下的偏误就是向零的偏误；另一方面，如果 β_1 为负，那么 $\widetilde{\beta}_1$ 向上的偏误就是向零的偏误。

遗漏变量的偏误：更一般的情形

在所估计的模型中含有多个回归元的情况下，推导遗漏变量偏误的符号则更为困难。我们必须记得，一个解释变量与误差之间存在相关，一般会导致所有 OLS 估计量都产生偏误。例如，假设总体模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \tag{3.49}$$

满足假定 MLR.1～MLR.4。但我们遗漏了变量 x_3，并估计了模型

$$\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1 + \widetilde{\beta}_2 x_2 \tag{3.50}$$

现在假设 x_2 和 x_3 无关，但 x_1 和 x_3 却相关。换句话说，x_1 与被遗漏的变量 x_3 相关，但 x_2 却与 x_3 无关。我们不禁要想，基于上一小节的推导，$\widetilde{\beta}_1$

92 可能是有偏的，但因为 x_2 与 x_3 无关，所以 $\tilde{\beta}_2$ 可能是无偏的。不幸的是，一般并非如此：$\tilde{\beta}_1$ 和 $\tilde{\beta}_2$ 通常都是有偏误的。惟一例外的是，x_1 和 x_2 不相关。

即使在上述相当简单的模型中，也很难得到 $\tilde{\beta}_1$ 和 $\tilde{\beta}_2$ 偏误的方向。这是因为，x_1、x_2 和 x_3 可能会两两相关。不过，一种近似方法在实践中常常很有用：如果假定 x_1 和 x_2 无关，那我们就能像在总体和所估计模型中没有 x_2 一样去研究 $\tilde{\beta}_1$ 的偏误。实际上，当 x_1 和 x_3 无关时，可以证明

$$\mathrm{E}(\tilde{\beta}_1)=\beta_1+\beta_3\frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)x_{i3}}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}$$

这就像方程（3.46）一样，只是 β_3 取代了 β_2，x_3 取代了 x_2。因此，通过在表 3.2 中用 β_3 取代 β_2 和用 x_3 取代 x_2，就得到了 $\tilde{\beta}_1$ 的偏误。如果 $\beta_3>0$，而且 $\mathrm{Corr}(x_1, x_3)>0$，则 $\tilde{\beta}_1$ 的偏误为正。如此等等。

作为一个例子，假设在工资模型中增加 *exper*：

$$wage=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 abil+u$$

如果模型中漏掉了 *abil*，则即使假定 *exper* 与 *abil* 无关，$\tilde{\beta}_1$ 和 $\tilde{\beta}_2$ 的估计量也都是有偏的。由于我们主要感兴趣的是教育回报，所以如果能断定 $\tilde{\beta}_1$ 因遗漏了能力变量而产生向上或向下的偏误，那就太好了。如果没有进一步的假定，这个结论也不可能得到。作为一种近似，我们姑且假定，除了 *exper* 与 *abil* 无关外，*educ* 与 *exper* 也不相关。（实际上，它们多少有些负相关。）由于 $\beta_3>0$，而且 *educ* 与 *abil* 正相关，所以就像模型中没有 *exper* 一样，$\tilde{\beta}_1$ 将会向上偏误。

上例中所使用的逻辑，通常被认为是要在更复杂的模型中得到估计量偏误的一个可能的粗略指导。通常关心的是，一个特定的解释变量（比方说 x_1）与遗漏的关键因素之间的关系。严格地讲，虽然只有在所有其他解释变量都与 x_1 无关时忽略它们才是令人信服的做法，但这样做仍不失为一个有用的指导原则。

3.4 OLS 估计量的方差

93 除了知道 $\hat{\beta}_j$ 的集中趋势之外，我们还想度量其在样本分布中分散状况的一种度量，所以现在来看 OLS 估计量的方差。在发现方差之前，像在第 2 章中一样，我们增加一个同方差性假定。首先，通过施加误差的方差为常数的假定，可以简化公式。其次，在 3.5 节我们将看到，如果增加了同方差性假定，OLS 就具有一个重要的性质，即有效性。

在多元回归的框架中，同方差性表述如下。

假定 MLR.5（同方差性）

$$\mathrm{Var}(u \mid x_1, \cdots, x_k) = \sigma^2$$

假定 MLR.5 意味着以解释变量为条件的误差项 u 的方差，不管解释变量出现怎样的组合，都是一样的。如果这个假定不成立，那么模型就像在两变量情形中一样表现出异方差性。

在方程

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

中，同方差性要求不可观测的误差 u 的方差不依赖于受教育程度、工作经历和现有任期水平。即

$$\mathrm{Var}(u \mid educ, exper, tenure) = \sigma^2$$

如果这个方差随着这三个解释变量中任何一个的变化而变化，就出现异方差性。

假定 MLR.1～MLR.5 一起被称为（横截面回归的）**高斯-马尔科夫**（Gauss-Markov）假定。迄今为止，我们对假定的表述都只适用于随机抽样的横截面分析。如我们将会见到的那样，对于时间序列分析或其他诸如综列数据分析之类的情形，高斯-马尔科夫假定的表述尽管也有许多相似之处，但更加困难。

在接下来的讨论中，我们将用 $\boldsymbol{x}$ 表示所有自变量（x_1，x_2，…，x_k）的集合。于是，在工资以 *educ*，*exper* 和 *tenure* 作为自变量而进行的回归中，$\boldsymbol{x}$＝（*educ*，*exper*，*tenure*）。现在我们可以将假定 MLR.3 写成

$$\mathrm{E}(y \mid \boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

而假定 MLR.5 也就等同于 $\mathrm{Var}(y \mid \boldsymbol{x}) = \sigma^2$。用这种方式表达这两个假定，清楚地说明了假定 MLR.5 和假定 MLR.3 如何大为不同。假定 MLR.3 是说，给定 $\boldsymbol{x}$，y 的期望值对参数为线性，这显然取决于$(x_1, x_2, \cdots, x_k)$。而假定 MLR.5 则表明，给定 $\boldsymbol{x}$，y 的方差并不取决于自变量的值。

我们现在可以得到 $\hat{\beta}_j$ 的方差了，并再次取决于自变量的样本值。其证明见本章的附录。

94

定理 3.2（OLS 斜率估计量的抽样方差）

在假定 MLR.1～MLR.5 之下，以自变量的样本值为条件，对所有的 $j = 1, 2, \cdots, k$，都有

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\mathrm{SST}_j(1 - R_j^2)} \tag{3.51}$$

式中，$\mathrm{SST}_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ 为 x_j 的总样本变异；R_j^2 为将 x_j 对所有其他自变量（并包括一个截距项）进行回归所得到的 R^2。

在我们更详尽地研究方程（3.51）之前，重要的是知道在得到这个公式的过程中，用到了所有的高斯-马尔科夫假定。虽然我们在断言 OLS 是无偏

估计时不需要同方差性假定，但要使式（3.51）成立，则必然需要这个同方差性假定。

$\mathrm{Var}(\hat{\beta}_j)$的大小在实践中也很重要。方差越大，则意味着估计量越不精确，也就是置信区间越大和假设检验越不准确（如我们将在第4章所看到的那样）。在接下来的小节中，我们讨论构成式（3.51）的要素。

OLS方差的成分：多重共线性

方程（3.51）表明，$\hat{\beta}_j$ 的方差取决于三个因素：σ^2，SST_j 和 R_j^2。记住，下标 j 只是表示自变量中的任意一个（如受教育程度或贫穷率等）。现在依次讨论影响 $\hat{\beta}_j$ 的每个因素。

误差方差 σ^2。从方程（3.51）来看，越大的 σ^2 意味着OLS估计量的方差就越大。对此根本不应该感到惊奇：方程中的“噪音”越多（σ^2 越大），使得估计任何一个自变量对 y 的局部影响都越困难，这将通过OLS斜率估计量的较大方差反映出来。由于 σ^2 是总体的一个特征，所以它与样本容量无关。它是式（3.51）中未知的一个组成部分。我们以后将看到如何得到 σ^2 的一个无偏估计量。

对于一个给定的因变量 y，确实只有一个办法减少误差方差，那就是在方程中增加更多的解释变量（将某些因素从误差项中取出来）。这样做不仅不一定可能，而且出于本章后面将要讨论的原因，还不一定总能令人满意。

x_j 的总样本变异 SST_j。从方程（3.51）来看，x_j 的总变异越大，$\mathrm{Var}(\hat{\beta}_j)$ 越小。因此，若所有其他条件不变，就估计 β_j 而言，我们希望 x_j 的样本方差越大越好。我们在第2章的简单回归情形中已经发现了这一点。尽管我们几乎不能选择自变量的样本值，但还是有一种办法来提高每个自变量的样本变异：扩大样本容量。实际上，当从总体中随机抽样时，随着样本容量越来越大，SST_j 将无限制地递增。这是方差中系统地取决于样本容量的部分。

95 若 SST_j 很小，$\mathrm{Var}(\hat{\beta}_j)$ 会变得很大。但小的 SST_j 并不违背假定MLR.4。从计算上讲，随着 SST_j 趋近于零，$\mathrm{Var}(\hat{\beta}_j)$ 可能趋于无穷大。但 x_j 无样本变异的这种极端情形（$\mathrm{SST}_j=0$）却是假定MLR.4所不允许的。

自变量之间的线性关系 R_j^2。方程（3.51）中的 R_j^2 一项，是这三个因素中最难理解的部分。由于在简单回归中只有一个自变量，所以不会出现这一项。重要的是要看出这个 R^2 不同于 y 对 x_1，x_2，…，x_k 回归所得到的 R^2：得到 R_j^2 的回归只涉及原模型中的自变量，其中 x_j 是作为因变量而出现的。

首先考虑 $k=2$ 的情形：$y=\beta_0+\beta_1x_1+\beta_2x_2+u$。当$\mathrm{Var}(\hat{\beta}_1)=\sigma^2/\mathrm{SST}_1(1-R_1^2)$时，$R_1^2$ 是 x_1 对 x_2（含截距）进行简单回归所得到的 R^2。由于 R^2 度量了拟合优度，所以当 R_1^2 值接近于1时，则表明在这个样本中，x_2 解释了 x_1 的大部分变动。这就意味着 x_1 和 x_2 高度相关。

随着 R_1^2 向1逐渐增加，$\mathrm{Var}(\hat{\beta}_1)$则变得越来越大。因此，$x_1$ 和 x_2 之间

线性关系的程度越高，OLS斜率估计值的方差就越大。（对 $\hat{\beta}_2$ 也可作类似论断。）从图3.1可以看出 $\mathrm{Var}(\hat{\beta}_1)$ 与从 x_1 对 x_2 进行简单回归中所得到的 R^2 之间的关系。

96

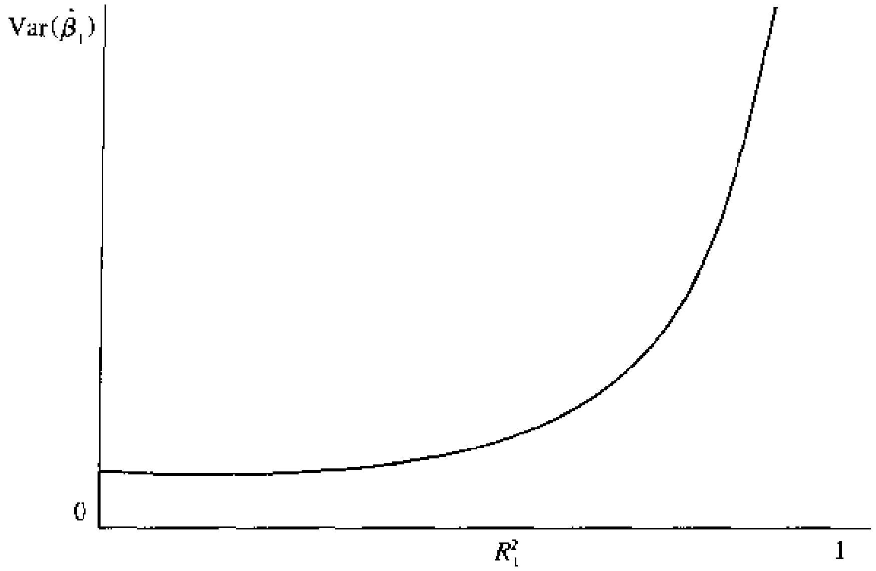


图3.1　作为 R_1^2 的一个函数的 $\mathrm{Var}(\hat{\beta}_1)$

一般情况下，R_j^2 是 x_j 总变异中可由方程中出现的**其他**自变量解释的部分。对于给定的 σ^2 和 SST_j，最小的 $\mathrm{Var}(\hat{\beta}_1)$ 在 $R_j^2=0$ 时得到，这种情况当且仅当 x_j 与其他**每个**自变量的样本相关均为零时才会发生。虽然这是估计 β_j 的最佳情形，却很难遇到。

另一个极端情形 $R_j^2=1$ 被假定MLR.4所排除，因为 $R_j^2=1$ 意味着 x_j **恰好**是回归中某些自变量的线性组合。还有一种更重要的情形是 R_j^2 “接近”于1的情况。我们从方程（3.51）和图3.1看到，这会导致 Var（$\hat{\beta}_j$）很大：随着 $R_j^2\to 1$，$\mathrm{Var}(\hat{\beta}_j)\to\infty$。两个或多个自变量之间高度（但不完全）相关被称为**多重共线性**（multicollinearity）。

在进一步讨论多重共线性问题之前，重要的是先弄清楚一点：R_j^2 接近于1并不违背假定MLR.4。

既然多重共线性没有违背任何一个假定，那么多重共线性的“问题”实际上就没有很好地定义。当我们说在 R_j^2 “接近”于1的情况下估计 β_j 可能会导致多重共线性时，把“接近”一词放在引号中，因为我们不能给出一个绝对的数字来断定什么情况下多重共线性会成为一个问题。比如，$R_j^2=0.9$ 意味着 x_j 样本变异中的90%都可由回归模型中的其他自变量来解释。毋庸置疑，这意味着与其他自变量之间存在着很强的线性关系。但这是否就能说明，因为 $\mathrm{Var}(\hat{\beta}_j)$ 太大，所以是无用的，却仍取决于 σ^2 和 SST_j。像我们在第4章将了解到的那样，就统计推断而言，最终的问题是 $\hat{\beta}_j$ 与其标准差相比有多大。

正如很大的 R_j^2 可以导致很大的 $\mathrm{Var}(\hat{\beta}_j)$ 一样，很小的 SST_j 也能导致很

大的 Var（$\hat{\beta}_j$）。因此，小样本容量也能导致很大的抽样方差。对样本中自变量间出现高度相关的担心，实际上无异于对小样本容量的担心：二者都会提高 Var（$\hat{\beta}_j$）。威斯康星大学著名的计量经济学家阿瑟·戈德伯格（Arthur Goldberger）针对计量经济学家对多重共线性的着迷情况，（略带挖苦意味地）创造了**微数缺测性**（micronumerosity）这个术语，并定义为“小样本容量的问题”。[对多重共线性和微数缺测性颇具吸引力的探讨，可参见戈德伯格（Goldberger，1991）。]

97 尽管不能清楚地界定多重共线性问题，但有一点是清楚的：在所有其他条件都不变的情况下，就估计 β_j 来说，x_j 与其他自变量之间越不相关越好。这一观察常常会引起对如何“解决”多重共线性问题的讨论。在社会科学中，我们通常都是被动的数据搜集者，除了搜集更多的数据外，没有什么好办法能减小无偏估计量的方差。对于一个给定的数据集，可以试着从模型中去掉一些其他自变量，以努力消除多重共线性。不幸的是，如我们在 3.3 节所见，去掉总体模型中的一个变量就会导致偏误。

可能此时举一个例子有助于澄清考虑多重共线性时出现的某些问题。假设我们想知道学校各种支出项目对学生表现的影响。对教师薪水、辅导材料和体育运动等方面的支出可能高度相关：富裕的学校在各方面的支出通常都高些，而贫穷的学校则在各方面的支出常常都低些。不必惊奇，如果一种支出项目的变异几乎全部能由其他支出项目的变异来解释（这会导致每个支出变量的 R_j^2 都很高），那么，就很难估计出某特定项目对学生表现的影响。虽然通过搜集更多的数据可以削弱这种多重共线性，但在某种意义上，我们已给自己强加了一个问题：我们正在分析一个可能因相对数据来说过于深奥而无法准确回答的问题。通过改变分析的范围并将所有的支出项目合并在一起，或许能做得更好，因为我们不再试图去估计每个孤立项目的局部影响。

另外一个重要的问题是，虽然某些自变量之间高度相关，但对模型中其他参数的估计效果而言，可能并不重要。比如，考虑一个含有三个自变量的模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

其中 x_2 和 x_3 高度相关。于是 $\mathrm{Var}(\hat{\beta}_2)$ 和 $\mathrm{Var}(\hat{\beta}_3)$ 都很大，但 x_2 和 x_3 之间的相关程度对 $\mathrm{Var}(\hat{\beta}_1)$ 没有直接影响。实际上，如果 x_1 与 x_2，x_3 无关，则无论 x_2 和 x_3 如何相关，都有 $R_1^2 = 0$ 和 $\mathrm{Var}(\hat{\beta}_1) = \sigma^2/\mathrm{SST}_1$。如果我们所关心的参数是 β_1，那么，我们实在不必关心 x_2 和 x_3 之间的相关程度。

问题 3.4

假设你提出一个模型，用出勤来解释期末考试成绩。因此，因变量是期末考试分数，而关键的解释变量是出勤次数。为了控制学生在课堂以外的能力和努力，你还在解释变量中增加了累积 *GPA*、*SAT* 分数和对高中表现的某种度量。有人说：“由于累积 *GPA*，*SAT* 分数和高中表现指标之间很可能高度共线性，所以你不要指望能从中得到什么结果。”你应该怎样回应呢？

前面这个观察之所以重要，是因为经济学家常常为了分离出某特定变量的因果性效应，而在模型中包括许多控制因素。比如，在考虑贷款许可率与邻居中少数民族所占比例之间的关系时，我们可能因为要得到有关歧视的因果性结论，而需要引入诸如平均收入、平均住房价格、信誉指标等变量。收入、住房价格和信誉彼此之间通常都高度相关，但这些变量之间的高度相关并没有使得对歧视效应的决定更为困难。

误设模型中的方差

在一个回归模型中是否包括一个特定变量的决策，可以通过分析偏误和方差之间的替换关系作出。在3.3节中，我们推导了当真实模型包含两个解释变量时，漏掉一个相关变量所产生的偏误。接着，我们比较了一些OLS估计量的方差，来继续分析这个模型。

将满足高斯-马尔科夫假定的真实总体模型写成

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u$$

98 我们考虑 β_1 的两个估计量。估计量 $\hat{\beta}_1$ 来自多元回归

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2 \tag{3.52}$$

换句话说，在回归模型中包括了 x_1 和 x_2。估计量 $\widetilde{\beta}_1$ 是在模型中漏掉 x_2 并将 y 对 x_1 进行如下简单回归所得到的：

$$\widetilde{y}=\widetilde{\beta}_0+\widetilde{\beta}_1x_1 \tag{3.53}$$

当 $\beta_2\neq0$ 时，方程就从模型中排除了一个有关变量，而且如我们在3.3节所见，除非 x_1 和 x_2 不相关，否则 $\widetilde{\beta}_1$ 就有偏误。另一方面，无论 β_2 等于多少，包括 $\beta_2=0$，$\hat{\beta}_1$ 都是 β_1 的无偏估计。于是，如果只以偏误为准则，那么 $\hat{\beta}_1$ 就比 $\widetilde{\beta}_1$ 好。

在考虑方差以后，$\hat{\beta}_1$ 总比 $\widetilde{\beta}_1$ 好的结论就不能继续成立。根据样本中 x_1 和 x_2 的值，我们从式（3.51）得到

$$\mathrm{Var}(\hat{\beta}_1)=\sigma^2/[\mathrm{SST}_1(1-R_1^2)] \tag{3.54}$$

式中，SST_1 为 x_1 的总变异；R_1^2 为 x_1 对 x_2 进行简单回归所得到的 R^2。而且，将第2章对两变量给出的证明稍加修改，即表明

$$\mathrm{Var}(\widetilde{\beta}_1)=\sigma^2/\mathrm{SST}_1 \tag{3.55}$$

式（3.55）与式（3.54）的比较表明，除非样本 x_1 和 x_2 不相关（在这种情况下，$\widetilde{\beta}_1$ 和 $\hat{\beta}_1$ 就是同一个估计量），否则 $\mathrm{Var}(\widetilde{\beta}_1)$ 总比 $\mathrm{Var}(\hat{\beta}_1)$ 小。假定 x_1 和 x_2 不相关，我们可以得到如下结论：

1. 当 $\beta_2\neq0$ 时，$\widetilde{\beta}_1$ 是有偏的，$\hat{\beta}_1$ 是无偏的，而且 $\mathrm{Var}(\widetilde{\beta}_1)<\mathrm{Var}(\hat{\beta}_1)$。
2. 当 $\beta_2=0$ 时，$\widetilde{\beta}_1$ 和 $\hat{\beta}_1$ 都是无偏的，而且 $\mathrm{Var}(\widetilde{\beta}_1)<\mathrm{Var}(\hat{\beta}_1)$。

从第二个结论看，如果$\beta_2=0$，显然$\widetilde{\beta}_1$更好。从直觉上讲，如果x_2对y没有局部影响，那么将它放在模型中，就只会加剧多重共线性的问题，从而导致β_1的估计量效率较低。在模型中包括了一个无关变量的代价是β_1的估计量方差较高。

$\beta_2\neq0$的情况就更困难些。不把x_2放到模型中，将导致β_1的估计量有偏误。按传统，计量经济学家曾建议，将因漏掉x_2而导致偏误的可能性大小与方差的降低（以R_1^2的大小来概括）相比较，以决定是否应该包括x_2。但当$\beta_2\neq0$时，有两个有利的原因让我们在模型中包括x_2。其中最重要的一个是，$\widetilde{\beta}_1$中的偏误不会随着样本容量的扩大而缩减；实际上，偏误不一定有任何原型。因此，我们可以有用地认为，偏误对任何样本容量都大致相等。另一方面，随着n逐渐变大，$\mathrm{Var}(\widetilde{\beta}_1)$和$\mathrm{Var}(\hat{\beta}_1)$都逐渐缩小至零，这意味着，随着样本容量逐渐变大，因增加x_2所导致的多重共线性就会变得没有那么重要。在大样本情况下，我们将更喜欢$\hat{\beta}_1$。

99 喜欢的另一个原因就更难以捉摸。式（3.55）中的方差公式取决于样本中x_{i1}和x_{i2}的值，这就为$\widetilde{\beta}_1$提供了最好的条件。当$\beta_2\neq0$时，$\widetilde{\beta}_1$的方差只取决于x_1，并且比式（3.55）中的方差更大。从直觉来看，当$\beta_2\neq0$而模型中又没有包括x_2时，误差方差因为误差有效地包含了部分x_2而提高。但式（3.55）还是因为将两个回归元都视为非随机变量而提高了误差方差。对取决于哪个自变量的充分讨论将使我们误入歧途，知道用式（3.55）来度量的准确性太过笼统就足够了。

估计σ^2：OLS估计量的标准误

现在来证明如何选择σ^2的一个无偏估计量，使我们能得到$\mathrm{Var}(\hat{\beta}_j)$的无偏估计量。

因为$\sigma^2=\mathrm{E}(u^2)$，故σ^2的无偏“估计量”就是误差平方的简单平均：$n^{-1}\sum_{i=1}^{n}u_i^2$。不幸的是，它因我们观测不到$u_i$而算不上一个真正的估计量。不过，回想误差可写成$u_i=y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\cdots-\beta_kx_{ik}$，所以观测不到$u_i$的原因是我们不知道$\beta_j$。当我们将每个$\beta_j$都用其OLS估计量取代后，就得到OLS残差

$$\hat{u}_i=y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1}-\hat{\beta}_2x_{i2}-\cdots-\hat{\beta}_kx_{ik}$$

看来用$\hat{u}_i$取代u_i来估计σ^2自然而然。在简单回归情形中，我们看到，这将导致一个有偏估计量。在一般多元回归情形中，σ^2的无偏估计量是

$$\hat{\sigma}^2=\left(\sum_{i=1}^{n}\hat{u}_i^2\right)/(n-k-1)\equiv\mathrm{SSR}/(n-k-1) \tag{3.56}$$

我们在$k=1$的简单回归中已经遇到了这个估计量。

式（3.56）中$n-k-1$这一项，是含有n个观测和k个自变量的一般

OLS问题的**自由度**（***df***）(degrees of freedom)。由于在一个含有 k 个自变量和一个截距项的回归模型中有 $k+1$ 个参数，所以可以得到

$$df = n-(k+1) = \text{观测个数} - \text{被估计参数的个数} \tag{3.57}$$

这是计算特定应用中的自由度最容易的方法：从观测个数中减去包括截距在内的参数个数。(在极少的情况下不估计截距，参数的个数减少一个。)

从技术上讲，式（3.56）中除以 $n-k-1$ 是因为残差平方和的期望值为E（SSR）=（$n-k-1$）σ^2 的事实。从直觉来看，通过回到OLS估计量的一阶条件，我们就能弄明白为什么需要对自由度进行调整。这些可以写成
100 $\sum_{i=1}^{n}\hat{u}_i = 0$ 和 $\sum_{i=1}^{n}x_{ij}\hat{u}_i = 0(j = 1,2,\cdots,k)$。因此，在得到OLS估计值时，对OLS残差施加了 $k+1$ 个限制。这意味着，给定残差中的 $n-$（$k+1$）个，余下的 $k+1$ 个是已知的：残差中只有 $n-$（$k+1$）个自由度。(与之相对比的是，**误差** u_i 在样本中有 n 个自由度。)

为便于参考，我们将这个讨论放在定理3.3中。我们在第二章曾针对简单回归分析的情况证明了这个定理（参见定理2.3)。(一般化的证明需要矩阵代数方面的知识，并在附录E中给出。)

定理 3.3（σ^2 的无偏估计）

在高斯-马尔科夫假定MLR.1～MLR.5下，E($\hat{\sigma}^2$) = σ^2。

$\hat{\sigma}_2$ 正的平方根 $\hat{\sigma}$ 被称为**回归标准误**或SER。SER是误差项之标准离差的估计量。这个估计值通常由回归软件包报告，尽管不同的软件包有不同的说法。(除了SER外，$\hat{\sigma}$ 也被称为**估计值的标准误**或**均方根误**。)

注意到，(对于给定的样本）在方程中增加另一个自变量时，$\hat{\sigma}$ 可能减小或增大。这是因为，当增加另一个解释变量时，在SSR肯定下降的同时，自由度也减少一个。因为SSR在分子中，而 df 在分母中，所以我们事先并不知道哪个作用会占主导地位。

为了能在第4章构造置信区间并进行检验，我们还需要估计 $\hat{\beta}_j$ 的**标准离差**，也就是方差的平方根：

$$\mathrm{sd}(\hat{\beta}_j) = \frac{\sigma}{[\mathrm{SST}_j(1-R_j^2)]^{1/2}}$$

由于 σ 未知，所以用其估计量 $\hat{\sigma}$ 来取代。这就给出了 $\hat{\beta}_j$ 的标准离差：

$$\mathrm{se}(\hat{\beta}_j) = \frac{\hat{\sigma}}{[\mathrm{SST}_j(1-R_j^2)]^{1/2}} \tag{3.58}$$

恰如对任何样本都能得到OLS估计值一样，标准误也能同样得到。由于se（$\hat{\beta}_j$）取决于 $\hat{\sigma}$，所以标准误有一个抽样分布，这在第4章将发挥一种作用。

对于标准误，我们应该强调一点。因为式（3.58）直接从方差公式（3.51）得到，而式（3.51）又依赖于同方差性假定MLR.5，所以，如果误差表现出异方差性，式（3.58）中的标准误公式就**不是**sd（$\hat{\beta}_j$）的一个可靠

估计量。于是，异方差性的出现，尽管不会导致 $\hat{\beta}_j$ 的偏误，却能导致 Var($\hat{\beta}_j$) 的常用公式产生偏误，从而使标准误无效。这一点很重要，因为任何回归软件包都计算式（3.58）作为对每个系数有问题的标准误（对于截距则多少有些不同）。如果我们怀疑存在异方差性，那么“通常”的 OLS 标准误就会无效，而且应该采取某种正确的行动。我们在第 8 章将看到，有哪些方法可用于处理异方差性。

3.5 OLS 的有效性：高斯-马尔科夫定理

101 在本节中，我们陈述并讨论重要的高斯-马尔科夫定理，它向我们说明了，为什么要使用 OLS 方法，而不是使用一系列其他与之相媲美的估计量。我们已经知道了使用 OLS 的依据：在假定 MLR.1～MLR.4 下，OLS 是无偏的。但在这些假定之下，β_j 还有许多其他的无偏估计量（参见习题 3.12）。还可能有其他的无偏估计量的方差比 OLS 估计量的方差更小吗？

如果适当限制这些各不相让的估计量的范围，就能证明 OLS 是这些估计量中最好的一个。具体而言，我们将证明，在假定 MLR.1～MLR.5 下，β_j 的 OLS 估计量 $\hat{\beta}_j$ 是**最优线性无偏估计量**（best linear unbiased estimator，BLUE）。为了表述这个定理，我们需要理解“BLUE”这个首字母缩拼词每个字母的含义。首先，我们知道什么是一个估计量：它是一个可应用于任何一个数据样本，并产生一个估计值的规则。我们也知道什么是一个无偏估计量：在当前背景下，如果 β_j 的一个估计量（比方说 $\tilde{\beta}_j$）对任意的 β_0，β_1，…，β_k 都有 E（$\tilde{\beta}_j$）$=\beta_j$，那么它就是 β_j 的一个无偏估计量。

“线性”一词的含义是什么呢？β_j 的一个估计量 $\tilde{\beta}_j$ 是线性的充分必要条件是，它能表示成因变量数据的一个线性函数：

$$\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i \tag{3.59}$$

式中，每个 w_{ij} 都可以是所有自变量样本值的一个函数。如从方程（3.22）所见，OLS 估计量就是线性的。

最后，我们如何定义“最优”呢？对于现在这个定理来说，最优被定义为**最小方差**。给定两个无偏估计量，喜欢方差最小的那个是合乎逻辑的（参见附录 C）。

现在，令 $\hat{\beta}_0$，$\hat{\beta}_1$，…，$\hat{\beta}_k$ 表示在假定 MLR.1～MLR.5 下模型（3.31）中的 OLS 估计量。高斯-马尔科夫定理是说，对任何一个线性无偏估计量 $\tilde{\beta}_j$，都有 Var($\hat{\beta}_j$)$\leqslant$Var($\tilde{\beta}_j$)，而且通常都取严格不等号。换句话说，在一群线性无偏估计量中，OLS 具有最小的方差（在五个高斯-马尔科夫假定之下）。实际上，这个定理的内涵还不止于此。如果我们想估计 β_j 的任何一个线性方程，那么，OLS 估计量的对应线性组合在所有的线性无偏估计量中

也体现了最小方差。我们以一个定理来做个总结，这个定理在附录 3A 中证明。

定理 3.4（高斯-马尔科夫定理）

在假定 MLR.1～MLR.5 下，$\hat{\beta}_0$，$\hat{\beta}_1$，…，$\hat{\beta}_k$ 分别是 β_0，β_1，…，β_k 的最优线性无偏估计量（BLUEs）。

正是因为这个定理，假定 MLR.1～MLR.5 才被称为（横截面数据分析的）高斯-马尔科夫假定。

高斯-马尔科夫定理的重要性在于，当这个标准假定集成立时，我们不 102 需要再去寻找其他形如式（3.59）的无偏估计量：没有一个会比 OLS 更好。换言之，如果如果有人向我们提出一个线性无偏估计量，那我们就知道，这个估计量的方差至少和 OLS 估计量的方差一样大；无须另做计算来说明这一点。

就我们的目的而言，定理 3.4 辨明了估计多元回归模型时使用 OLS 的合理性。如果高斯-马尔科夫假定中的任何一个不成立，那么这个定理也就不再成立。我们已经知道，零条件均值的假定（假定 MLR.3）不成立会导致 OLS 产生偏误，所以定理 3.4 也不成立。我们还知道，异方差性（假定 MLR.5 不成立）虽不致使 OLS 有偏，但它在线性无偏估计量中不再具有最小方差。在第 8 章，我们分析了一个在出现异方差性时能比 OLS 有所改进的估计量。

小　结

1. 多元回归模型能使我们有效地在保持其他因素不变的情况下，考察一个特定的自变量对因变量的影响。显然可见，它允许自变量之间存在相关。

2. 尽管这个模型对参数为线性，但通过适当选取因变量和自变量，它也可用于对非线性关系加以线性化。

3. 普通最小二乘法很容易应用于多元回归模型。每个斜率参数都度量了，在保持所有其他自变量不变的条件下，相应的自变量对因变量的局部影响（或偏效应）。

4. R^2是因变量样本变异中能被自变量解释的部分，并用来度量回归的拟合优度。重要的是，不要在评价计量模型时对 R^2 的值寄予太大希望。

5. 在前四个高斯-马尔科夫假定（假定 MLR.1～MLR.4）下，OLS 估计量是无偏的。这意味着，在模型中包括一个无关变量对截距和其他斜率估计量的无偏性没有任何影响。另一方面，遗漏一个相关变量则导致 OLS 产生偏误。在许多情形下，偏误的方向都是可以确定的。

6. 在五个高斯-马尔科夫假定下，OLS 斜率估计量的方差由 $\mathrm{Var}(\hat{\beta}_j)$ =

$\sigma^2/[\mathrm{SST}_j(1-R_j^2)]$给出。随着误差方差 σ^2 的增加，$\mathrm{Var}(\hat{\beta}_j)$也会增加，但随着 x_j 的样本变异 SST_j 的提高，$\mathrm{Var}(\hat{\beta}_j)$则下降。$R_j^2$ 度量的是，x_j 与其他解释变量之间共线性的大小。当 R_j^2 接近于1时，$\mathrm{Var}(\hat{\beta}_j)$趋于无限。

7. 在方程中增加一个无关变量，通常都会因多重共线性而提高其余的OLS估计量的方差。

8. 在高斯-马尔科夫假定（假定MLR.1～MLR.5）下，OLS估计量是最优线性无偏估计量（BLUE）。

关键术语

103 最优线性无偏估计量（BLUE）
向零偏误
其他条件不变
自由度（*df*）
干扰
向下的偏误
内生解释变量
误差项
排除一个有关变量
外生解释变量
解释平方和（SSE）
一阶条件
高斯-马尔科夫假定
高斯-马尔科夫定理
包含一个无关变量
截距
微数缺测性
误设分析
多重共线性
多元线性回归模型
多元回归分析
变量遗漏偏误
OLS截距估计值
OLS回归线
OLS斜率估计值
普通最小二乘
模型过度设定
局部影响（偏效应）
完全共线性
总体模型
残差
残差平方和
样本回归函数（SRF）
斜率参数
$\hat{\beta}_j$ 的标准差
$\hat{\beta}_j$ 的标准误
回归标准误（SER）
残差平方和（SSR）
总平方和（SST）
真实模型
模型设定不足
向上偏误

习　题

3.1 利用GPA2.RAW中有关4 137名大学生的数据，用OLS估计了如下方程：

$$\widehat{colgpa} = 1.392 - 0.013\,5\,hsperc + 0.001\,48\,sat$$
$$n = 4\,137, R^2 = 0.273$$

式中，$colgpa$ 以四分制度量；$hsperc$ 为在高中班上名次的百分位数（比方说，$hsperc = 5$，就意味着位于班上前5%之列）；sat 为在学生能力测验中数学和语言的综合成绩。

(i) 为什么 $hsperc$ 的系数为负是合乎情理的？

(ii) 当 $hsperc = 20$ 和 $sat = 1\,050$ 时，大学 *GPA* 的预测值是多少？

(iii) 假设两个在高中班上具有同样百分位数的高中毕业生A和B，但A学生的 *SAT* 分数要高出140分（在样本中相当于1倍的标准离差），那么，预计这两个学生的大学 *GPA* 会相差多少？这个差距大吗？

(iv) 保持 $hsperc$ 不变，*SAT* 的分数相差多少，才能导致预测的 $colgpa$ 相差0.50或四分制的半分？评论你的结论。

104 3.2 用WAGE2.RAW中有关男工人的数据估计了如下方程：

$$\widehat{educ} = 10.36 - 0.094\,sibs + 0.131\,meduc + 0.210\,feduc$$
$$n = 722, R^2 = 0.214$$

式中，$educ$ 为受教育年数；$sibs$ 为兄弟姐妹的个数；$meduc$ 为母亲受教育的年数；$feduc$ 为父亲受教育的年数。

(i) $sibs$ 是否具有预期的影响？请给出解释。保持 $meduc$ 和 $feduc$ 不变，为了使预测的受教育水平减少一年，需要 $sibs$ 增加多少？（不要求答案为整数。）

(ii) 讨论对 $meduc$ 的系数的解释。

(iii) 假设一个男工人A没有兄弟姐妹，其父母都接受了12年的教育。另一个男工人B也没有兄弟姐妹，但其父母都接受了16年的教育。预计B和A所接受教育的年数差别为多少？

3.3 下面这个模型是比德尔和哈默什（Biddle and Hamermesh，1990）所用多元回归模型的一个简化版本，原模型研究睡眠时间和工作时间之间的取舍，并考查其他影响睡眠的因素：

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u$$

式中，$sleep$ 和 $totwrk$ 都以分钟/周为单位；$educ$ 和 age 以年为单位。（还可参见习题2.12。）

(i) 如果成年人为工作而放弃睡眠，β_1 的符号是什么？

(ii) 你认为 β_2 和 β_3 的符号应该是什么？

(iii) 利用SLEEP75.RAW中的数据，估计出来的方程是

$$\widehat{sleep} = 3\,638.25 - 0.148\,totwrk - 11.13\,educ + 2.20\,age$$
$$n = 706, R^2 = 0.113$$

如果有人一周多工作5个小时，预计 $sleep$ 会减少多少分钟？这是一个很大的取舍吗？

(iv) 讨论 $educ$ 的估计系数的符号和大小。

(v) 你能说 *totwrk*，*educ* 和 *age* 解释了变异中的大部分吗？其他还有什么因素可能影响花在睡眠上的时间？它们与 *totwrk* 可能相关吗？

3.4　刚从法学院毕业的学生的起薪中位数由下式决定：

$$\log(salary)=\beta_0+\beta_1 LSAT+\beta_2 GPA+\beta_3\log(libvol)+\beta_4\log(cost)+\beta_5 rank+u$$

式中，*LSAT** 为整个待毕业年级 *LSAT* 成绩的中位数；*GPA* 为该年级大学 *GPA* 的中位数；*libvol* 为法学院图书馆的藏书量；*cost* 为进入法学院每年的费用；*rank* 为法学院的排名（*rank* = 1 的法学院是最好的）。

105 (i) 解释为什么我们预期 $\beta_5\leqslant 0$。

(ii) 你预计其他斜率参数的符号如何？给出你的理由。

(iii) 使用 LAWSCH85.RAW 中的数据，估计出来的方程是

$$\log(\widehat{salary})=8.34+0.0047LSAT+0.248GPA+0.095\log(libvol)+0.038\log(cost)-0.0033rank$$
$$n=136, R^2=0.842$$

在其他条件不变的情况下，预计 *GPA* 中位数相差 1 分会导致薪水有多大的差别？（以百分比回答。）

(iv) 解释变量 log（*libvol*）的系数。

(v) 你是否认为应该进入一个排名更高的法学院？从预计的起薪来看，排名相差 20 位的价值有多大？

3.5　在一项调查大学 *GPA* 与在各种活动中所耗费时间之关系的研究中，你对几个学生分发了调查问卷。学生被问到他们每周在学习、睡觉、工作和闲暇这四种活动中各花多少小时。任何活动都被列为这四种活动之一，所以对每个学生来说，这四个活动的小时数之和都是 168。

(i) 在模型中 $GPA=\beta_0+\beta_1 study+\beta_2 sleep+\beta_3 work+\beta_4 leisure+u$ 中，保持 *sleep*，*work* 和 *leisure* 不变而改变 *study* 是否有意义？

(ii) 解释为什么这个模型违背了假定 MLR.4。

(iii) 你如何才能将这个模型重新表述，使得它的参数具有一个有用的解释，而又不违背假定 MLR.4。

3.6　考虑含有三个自变量的多元回归模型，并满足假定 MLR.1～MLR.4：

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+u$$

你对估计 x_1 和 x_2 的参数和感兴趣；称这个和为 $\theta_1=\beta_1+\beta_2$。证明 $\hat{\theta}_1=\hat{\beta}_1+\hat{\beta}_2$ 是 θ_1 的一个无偏估计量。

3.7　下面哪个因素会导致 OLS 估计量出现偏误？

(i) 异方差性。

(ii) 遗漏一个重要变量。

---

* LSAT 为美国的法学院入学考试（Law School Admission Test）的简称。——译者注

(iii) 模型中同时包含的两个自变量之间的样本相关系数达到0.95。

3.8 假设制造业中每个工人的平均生产力（*avgprod*）取决于培训的平均小时数（*avgtrain*）和工人的平均能力（*avgabil*）两个因素：

$$avgprod = \beta_0 + \beta_1 avgtrain + \beta_2 avgabil + u$$

106 假设这个方程满足高斯-马尔科夫假定。如果将培训津贴给了那些工人能力较差的企业，以致 *avgtrain* 和 *avgabil* 呈负相关，那么，将 *avgprod* 对 *avgtrain* 进行简单回归所得到的 $\tilde{\beta}_1$，可能出现什么样的偏误？

3.9 下面这个方程，用一个社区中污染量（*nox*，用来表示氧化亚氮）和每套住房的平均房间个数（*rooms*），来解释该社区内的平均住房价格：

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 rooms + u$$

(i) β_1 和 β_2 可能的符号是什么？对 β_1 该如何解释？试解释。

(ii) 为什么 *nox* ［更准确地说，是 log（*nox*）］和 *rooms* 可能负相关？如果是这样，将 log（*price*）对 log（*nox*）进行简单回归，会使 β_1 成为一个向上或向下偏误的估计量吗？

(iii) 使用 HPRICE2.RAW 中的数据，可估计如下方程：

$$\log(\widehat{price}) = 11.71 - 1.043\log(nox)$$
$$n = 506, R^2 = 0.264$$
$$\log(\widehat{price}) = 9.23 - 0.718\log(nox) + 0.306 rooms$$
$$n = 506, R^2 = 0.514$$

price 对 *nox* 的弹性的简单回归和多元回归估计值之间的关系是否与你的预期相符［给定你在第（ii）部分的答案］？这是否意味着，-0.718 就肯定比 -1.043 更接近于真正的弹性呢？

3.10 假设决定的总体模型是

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

而这个模型满足高斯-马尔科夫假定。但我们估计了漏掉 x_3 的模型。令 $\tilde{\beta}_0$，$\tilde{\beta}_1$ 和 $\tilde{\beta}_2$ 为 y 对 x_1 和 x_2 回归的 OLS 估计量。（给定样本中自变量的值）证明 $\tilde{\beta}_1$ 的期望值是

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n} \hat{r}_{i1} x_{i3}}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$$

式中，$\hat{r}_{i1}$为 x_1 对 x_2 回归所得到的 OLS 残差。［提示：$\tilde{\beta}_1$ 的公式来自方程(3.22)。在这个方程中加上$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$。经过一些计算之后，将 x_{i3}和 $\hat{r}_{i1}$作为非随机变量而取期望。］

3.11 下面这个方程表示各种税收比例对美国各县总体在随后就业增长方面的影响：

$$growth = \beta_0 + \beta_1 share_P + \beta_2 share_I + \beta_3 share_S + \text{其他因素}$$

式中，*growth* 为就业从 1980 年到 1990 年的变化百分比；$share_P$ 为总税收收益中财产税的比例；$share_I$ 为收入税税收收益的比例；$share_S$ 为销售税税收
107 收益的比例。所有这些变量都以 1980 年的货币度量。遗漏的比例 $share_F$ 包括收费和各种税收。根据定义，这四个比例之和为 1。其他因素将包括对教育、基础设施等支出（均以 1980 年货币度量）。

（i）我们为什么必须从方程中省略一个税收比例变量？

（ii）对 β_1 给出一个仔细的解释。

3.12 （i）在前四个高斯-马尔科夫假定之下，考虑简单回归模型 $y=\beta_0+\beta_1x_1+u$。对某个函数 $g(x)$，比如 $g(x)=x^2$ 或 $g(x)=\log(1+x^2)$，定义 $z_i=g(x_i)$。定义一个斜率估计量为

$$\tilde{\beta}_1=\frac{\sum_{i=1}^{n}(z_i-\bar{z})y_i}{\sum_{i=1}^{n}(z_i-\bar{z})x_i}$$

证明 $\tilde{\beta}_1$ 是线性无偏的。记住，在你的推导过程中，因为 $\mathrm{E}(u\mid x)=0$，所以你可以把 x_i 和 z_i 都看成非随机的。

（ii）增加同方差性假定 MLR.5，证明

$$\mathrm{Var}(\tilde{\beta}_1)=\frac{\sigma^2\left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]}{\left[\sum_{i=1}^{n}(z_i-\bar{z})x_i\right]^2}$$

（iii）在高斯-马尔科夫假定之下，直接证明 $\mathrm{Var}(\hat{\beta}_1)\leqslant\mathrm{Var}(\tilde{\beta}_1)$，其中 $\hat{\beta}_1$ 是 OLS 估计量。［提示：附录 B 中的柯西-施瓦兹不等式意味着

$$\left[n^{-1}\sum_{i=1}^{n}(z_i-\bar{z})(x_i-\bar{x})\right]^2\leqslant\left[n^{-1}\sum_{i=1}^{n}(z_i-\bar{z})^2\right]\left[n^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right];$$

注意我们可以将 $\bar{x}$ 从样本协方差中去掉。］

计算机习题

3.13 健康官员（和其他人）所关心的一个问题是，孕妇在怀孕期间吸烟对婴儿健康的影响。对婴儿健康的度量方法之一是婴儿出生时的体重，过低的出生体重会使婴儿有感染各种疾病的危险。由于除了吸烟之外，其他影响婴儿出生体重的因素可能与吸烟相关，所以我们应该考虑这些因素。比如，高收入通常会使母亲得到更好的产前照顾和更好的营养。表达这一点的一个方程是

$$bwght=\beta_0+\beta_1cigs+\beta_2faminc+u$$

（i）β_2 的符号最可能是什么？

（ii）你是否认为 *cigs* 和 *faminc* 可能相关？解释为什么可能是正或负相关。

108 （iii）现在利用 BWGHT.RAW 中的数据估计有和没有 *faminc* 的方程。

以方程的形式报告结论，包括样本容量和 R^2。讨论你的结论，主要看增加 *faminc* 是否会显著改变 *cigs* 对 *bwght* 的估计影响。

3.14　使用 HPRICE1.RAW 中的数据，估计如下模型：

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

式中，*price* 为以千美元为单位的住房价格。

(i) 以方程的形式写出结果。

(ii) 住房在保持面积不变而又增加一间卧室，估计其价格会提高多少？

(iii) 住房增加一间大小为 140 平方英尺的卧室，估计其价格会提高多少？将这个答案与你在第 (ii) 部分的答案相比较。

(iv) 价格的变异有多大比例能被平方英尺和卧室数解释？

(v) 样本中的第一套住房有 $sqrft = 2\ 438$ 和 $bdrms = 4$。从 OLS 回归线计算这套住房的预计销售价格。

(vi) 样本中第一套住房的实际销售价格是300 000美元（$price = 300$）。求出这套住房的残差，它是否表明购买者为这套住房支付了过低或过高的价格？

3.15　文件 CEOSAL2.RAW 包含了 177 位总经理的数据，并可用来考查企业业绩对 CEO 薪水的影响。

(i) 估计一个将年薪与企业销售量和市场价值相关的模型。让这个模型对每个自变量的变化都具有常弹性。以方程的形式写出结论。

(ii) 在第 (i) 部分的模型中增加 *profits*。为什么这个变量不能以对数形式进入模型？你会说这些企业业绩变量解释了 CEO 薪水变异中的大部分吗？

(iii) 在第 (ii) 部分的模型中增加 *ceoten*。保持其他条件不变，延长一年 CEO 任期，估计的百分比回报是什么？

(iv) 求出变量 log（*mktval*）和 *profits* 之间的样本相关系数。这些变量高度相关吗？这对 OLS 估计量有何含义？

3.16　本题中使用 ATTEND.RAW 中的数据。

(i) 求出变量 *atndrte*，*priGPA* 和 *ACT* 的最小、最大和平均值。

(ii) 估计模型

$$atndrte = \beta_0 + \beta_1 priGPA + \beta_2 ACT + u$$

并以方程的形式写出结论。对截距作出解释。它是否有一个有用的含义。

(iii)讨论估计的斜率系数。有没有什么令人吃惊之处？

(iv)如果 $priGPA = 3.65$ 和 $ACT = 20$，预计 *atndrte* 是多少？你对这个结论做何解释？样本中有没有一些学生具有这些解释变量的值？

109 (v)如果学生 A 具有 $priGPA = 3.1$ 和 $ACT = 21$，而学生 B 具有 $priGPA = 2.1$和 $ACT = 26$，他们在出勤率上的预期差异是多少？

3.17　通过对例 3.2 明确地进行“排除其他影响”的练习，证实对 OLS 估计值做“排除其他影响”的解释。这首先要求，将 *educ* 对 *exper* 和 *tenure*

进行回归，并保留残差 $\hat{r}_1$。然后将 log（*wage*）对 $\hat{r}_1$ 进行回归。将 $\hat{r}_1$ 的系数与在 log（*wage*）对 *educ*，*exper* 和 *tenure* 的回归中 *educ* 的系数相比较。

附录 3A

3A.1 对一阶条件方程（3.13）的推导

这个分析与简单回归情形极为类似。我们必须刻画问题

$$\min_{b_0, b_1, \cdots, b_k} \sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - \cdots - b_k x_{ik})^2$$

的解。对每个 b_j 求偏导（参见附录 A），代入解点数值，并令它们都等于零，即得到

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$-2\sum_{i=1}^{n}x_{ij}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0, \quad j = 1, \cdots, k$$

消掉 -2 就给出了式（3.13）中的一阶条件。

3A.2 对方程（3.22）的推导

为了推导方程（3.22），将 x_1 对 x_2，…，x_k 进行回归，用回归中 x_{i1} 的拟合值及其残差来表示 x_{i1}：对于 $i=1$，…，n，$x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$。现在将其代入式（3.13）中的第二个方程：

$$\sum_{i=1}^{n}(\hat{x}_{i1} + \hat{r}_{i1})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \qquad (3.60)$$

110 根据 OLS 残差 $\hat{u}_i$ 的定义，因为 $\hat{x}_{i1}$ 只是解释变量 x_{i2}，…，x_{ik} 的一个线性函数，于是 $\sum_{i=1}^{n}\hat{x}_{i1}\hat{u}_i = 0$。因此，式（3.60）可表示成

$$\sum_{i=1}^{n}\hat{r}_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0 \qquad (3.61)$$

因为 $\hat{r}_{i1}$ 是 x_1 对 x_2，…，x_k 进行回归的残差，所以对所有 $j=2$，…，k，都有 $\sum_{i=1}^{n} x_{ij}\hat{r}_{i1} = 0$。因此，式（3.61）就等价于 $\sum_{i=1}^{n}\hat{r}_{i1}(y_i - \hat{\beta}_1 x_{i1}) = 0$。最

后，我们使用 $\sum_{i=1}^{n}\hat{x}_{i1}\hat{r}_{i1}=0$ 的事实，它意味着 $\hat{\beta}_1$ 是 $\sum_{i=1}^{n}\hat{r}_{i1}(y_i-\hat{\beta}_1\hat{r}_{i1})=0$ 的解。现在，直接运算就得到式（3.22），当然要给定 $\sum_{i=1}^{n}\hat{r}_{i1}^2>0$；这从假定MLR.4得到肯定。

3A.3 对定理3.1的证明

我们对 $\hat{\beta}_1$ 证明定理3.1，其他斜率参数完全相同。（用矩阵给出的一个更简洁的证明，可参见附录E。）在假定MLR.4下，OLS估计量将存在，而且可以像在式（3.22）中那样写出 $\hat{\beta}_1$。在假定MLR.1下，我们可以像在式（3.32）中那样写出 y_i，并以此代入式（3.32）中的 y_i。于是，利用 $\sum_{i=1}^{n}\hat{r}_{i1}=0$，对所有 $j=2,\cdots,k$ 都成立的 $\sum_{i=1}^{n}x_{ij}\hat{r}_{i1}=0$ 和 $\sum_{i=1}^{n}x_{i1}\hat{r}_{i1}=\sum_{i=1}^{n}\hat{r}_{i1}^2$，我们得到

$$\hat{\beta}_1=\beta_1+\left(\sum_{i=1}^{n}\hat{r}_{i1}u_i\right)\Big/\left(\sum_{i=1}^{n}\hat{r}_{i1}^2\right) \tag{3.62}$$

现在，在假定MLR.2和MLR.4下，给定样本中所有自变量的值，每个 u_i 的期望值都是零。由于 $\hat{r}_{i1}$ 只是样本自变量的函数，于是

$$\begin{aligned}
&\mathrm{E}(\hat{\beta}_1|\boldsymbol{X})\\
&=\beta_1+\frac{\sum_{i=1}^{n}\hat{r}_{i1}\mathrm{E}(u_i|\boldsymbol{X})}{\sum_{i=1}^{n}\hat{r}_{i1}^2}\\
&=\beta_1+\frac{\sum_{i=1}^{n}\hat{r}_{i1}\cdot 0}{\sum_{i=1}^{n}\hat{r}_{i1}^2}\\
&=\beta_1
\end{aligned}$$

式中，$\boldsymbol{X}$ 为有关所有自变量的数据，给定所有观测 $i=1,\cdots,n$ 的 $x_{i1},\cdots,x_{ik}$，E（$\hat{\beta}_1|\boldsymbol{X}$）是 $\hat{\beta}_1$ 的期望值。证毕。

3A.4 对定理3.2的证明

同样，我们对 $j=1$ 证明这个定理。像在式（3.62）中那样写出 $\hat{\beta}_1$。现在，在假定MLR.5下，对所有的 $i=1,\cdots,n$，都有 Var（$u_i|\boldsymbol{X}$）$=\sigma^2$。在随机抽样条件下，u_i 虽以 $\boldsymbol{X}$ 为条件，但仍是独立的，而 $\boldsymbol{X}$ 为条件的 $\hat{r}_{i1}$ 又

是非随机的。因此

$$\begin{aligned}\mathrm{Var}(\hat\beta_1 \mid \boldsymbol{X}) &= \frac{\sum_{i=1}^{n}\hat r_{i1}^2\,\mathrm{Var}(u_i \mid \boldsymbol{X})}{\left(\sum_{i=1}^{n}\hat r_{i1}^2\right)^2}\\ &= \frac{\sum_{i=1}^{n}\hat r_{i1}^2\sigma^2}{\left(\sum_{i=1}^{n}\hat r_{i1}^2\right)^2}\\ &= \frac{\sigma^2}{\sum_{i=1}^{n}\hat r_{i1}^2}\end{aligned}$$

111 现在，因为 $\sum_{i=1}^{n}\hat r_{i1}^2$ 是 x_1 对 $x_2,\cdots,x_k$ 进行回归的残差平方和，所以 $\sum_{i=1}^{n}\hat r_{i1}^2 = \mathrm{SST}_1(1-R_1^2)$。证毕。

3A.5 对定理 3.4 的证明

我们现在来证明，对于其他任何一个 β_1 的线性无偏估计量 $\tilde\beta_1$，都有 $\mathrm{Var}(\tilde\beta_1)\geqslant\mathrm{Var}(\hat\beta_1)$，其中 $\hat\beta_1$ 是 OLS 估计量。只考虑 $j=1$ 仍不失其一般性。

对于如式（3.59）中那样的 $\tilde\beta_1$，可以代入 y_i，得到

$$\begin{aligned}\tilde\beta_1 = {}&\beta_0\sum_{i=1}^{n}w_{i1} + \beta_1\sum_{i=1}^{n}w_{i1}x_{i1} + \beta_2\sum_{i=1}^{n}w_{i1}x_{i2} + \cdots\\ &+ \beta_k\sum_{i=1}^{n}w_{i1}x_{ik} + \sum_{i=1}^{n}w_{i1}u_i\end{aligned}$$

现在，由于 w_{i1} 是 x_{ij} 的函数，而且对所有的 $i=1,\cdots,n$，在假定 MLR.3 下，都有 $\mathrm{E}(u_i \mid \boldsymbol{X})=0$，所以

$$\begin{aligned}\mathrm{E}(\tilde\beta_1 \mid \boldsymbol{X}) &= \beta_0\sum_{i=1}^{n}w_{i1} + \beta_1\sum_{i=1}^{n}w_{i1}x_{i1} + \beta_2\sum_{i=1}^{n}w_{i1}x_{i2} + \cdots\\ &\quad + \beta_k\sum_{i=1}^{n}w_{i1}x_{ik} + \sum_{i=1}^{n}w_{i1}\mathrm{E}(u_i \mid \boldsymbol{X})\\ &= \beta_0\sum_{i=1}^{n}w_{i1} + \beta_1\sum_{i=1}^{n}w_{i1}x_{i1} + \beta_2\sum_{i=1}^{n}w_{i1}x_{i2} + \cdots\\ &\quad + \beta_k\sum_{i=1}^{n}w_{i1}x_{ik}\end{aligned}$$

因此，要 $\mathrm{E}(\tilde\beta_1 \mid \boldsymbol{X})=\beta_1$ 对所有的参数值都成立，必须有

$$\sum_{i=1}^{n}w_{i1}=0,\ \sum_{i=1}^{n}w_{i1}x_{i1}=1,\ \sum_{i=1}^{n}w_{i1}x_{ij}=0,\ j=2,\cdots,k \quad (3.63)$$

现在，令 $\hat r_{i1}$ 为 x_{i1} 对 $x_{i2},\cdots,x_{ik}$ 进行回归所得到的残差。于是，从式(3.63)可以得到

$$\sum_{i=1}^{n} w_{i1}\hat{r}_{i1} = 1 \tag{3.64}$$

现在在假定MLR.1～MLR.5下，考虑Var($\tilde{\beta}_1|\boldsymbol{X}$)与Var($\hat{\beta}_1|\boldsymbol{X}$)的差：

$$= \sigma^2\sum_{i=1}^{n} w_{i1}^2 - \frac{\sigma^2}{\sum_{i=1}^{n} \hat{r}_{i1}^2} \tag{3.65}$$

因为式(3.64)，可以将式(3.65)中的差写成(不含σ^2)

$$\sum_{i=1}^{n} w_{i1}^2 - \frac{\left(\sum_{i=1}^{n} w_{i1}\hat{r}_{i1}\right)^2}{\sum_{i=1}^{n} \hat{r}_{i1}^2} \tag{3.66}$$

但式(3.66)不外是

$$\sum_{i=1}^{n} (w_{i1} - \hat{\gamma}_1\hat{r}_{i1})^2 \tag{3.67}$$

112

其中，$\hat{\gamma}_1 = \frac{\left(\sum_{i=1}^{n} w_{i1}\hat{r}_{i1}\right)^2}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$，通过将式(3.67)中的每一项都平方、求和及合并同类项，即可看出这一点。因为式(3.67)只是w_{i1}对$\hat{r}_{i1}$进行简单回归的残差平方和(记住：$\hat{r}_{i1}$的样本平均值为零)，所以式(3.67)一定非负。证毕。

第 4 章　多元回归分析：推断

113 本章继续对多元回归分析的讨论。现在转向对总体回归模型中的参数进行假设检验的问题。在总体误差服从正态分布这个新增假定下，我们从寻求OLS估计量的分布开始着手。4.2 节和 4.3 节涵盖了对单个参数的假设检验，而 4.4 节讨论的是如何检验一个涉及不止一个参数的假设。我们在 4.5 节对多重限制进行了检验，并特别关注是否应该从一个模型中省略掉一组自变量的问题。

4.1　OLS 估计量的抽样分布

到目前为止，我们已经建立了一系列假定，在这些假定之下，OLS是无偏的，而且还推导和讨论了由遗漏变量所导致的偏误。在 3.4 节中，我们在高斯-马尔科夫假定下得到了 OLS 估计量的方差。在 3.5 节，我们证明了在所有线性无偏估计量中，这个方差是最小的。

了解 OLS估计量的期望值和方差，有助于描述 OLS 估计量的精密度。但为了进行统计推断，我们需要知道 $\hat{\beta}_j$ 的不止这两个，还需要知道 $\hat{\beta}_j$ 的全部抽样分布。即使在高斯-马尔科夫假定下，$\hat{\beta}_j$ 的分布仍完全有可能具有任何形式。

当我们把样本中自变量的值视为既定时，显然 OLS 估计量的抽样分布就取决于其背后的误差分布。为了使得抽样分布易于掌握，我们现在假定，总体中不可观测的误差是正态分布的。我们称之为**正态性假定**（normality assumption）。

假定 MLR.6（正态性）

总体误差 u 独立于解释变量 x_1，x_2，…，x_k，而且服从均值为零和方差为 σ^2 的正态分布：$u \sim \text{Normal}(0,\sigma^2)$。

114 假定 MLR.6 是一个比我们前面任何一个假定都更强的假定。实际上，由于在假定 MLR.6 下 u 独立于 x_j，所以 $\text{E}(u \mid x_1, \cdots, x_k) = \text{E}(u) = 0$ 和 $\text{Var}(u \mid x_1, \cdots, x_k) = \text{Var}(u) = \sigma^2$。因此，如果作出假定 MLR.6，那我们就必然假定了 MLR.3 和 MLR.5。为了强调现在所做的假定比以前多，我们将使用假定 MLR.1～MLR.6 全套假定。

就横截面回归中的应用而言，假定 MLR.1～MLR.6 这六个假定被称为**经典线性模型（CLM）假定**［classical linear model（CLM）assumptions］。于是我们将这六个假定下的模型称为**经典线性模型**（classical linear model）。最好认为 CLM 假定包括了所有的高斯-马尔科夫假定，再加上误差正态分布的假定。

在 CLM 假定下，OLS 估计量 $\hat{\beta}_0$，$\hat{\beta}_1$，…，$\hat{\beta}_k$ 比在高斯-马尔科夫假定下具有更强的效率性质。可以证明，OLS 估计量**方差最小的无偏估计**（minimum variance unbiased estimator），即在所有的无偏估计中，OLS 具有最小的方差；不再需要把我们的比较限制在以 y_i 为线性的估计量内。CLM 假定下 OLS 的这个性质将在附录 E 中进一步讨论。

总结 CLM 总体假定的一种简捷方法是

$$y \mid \mathbf{x} \sim \text{Normal}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \sigma^2)$$

式中，$\mathbf{x}$ 为 $(x_1, x_2, \cdots, x_k)$ 的简写。就是说，以 $\mathbf{x}$ 为条件的 y，服从一个对 x_1，x_2，…，x_k 为线性均值并且方差为常数的正态分布。图 4.1 给出了只有一个自变量的情形。

推出误差服从正态分布的理由通常这样推理：由于 u 是影响着 y 而又观测不到的许多因素之和，所以我们可借助于中心极限定理（参见附录 C）断定 u 具有近似正态分布。虽然这种论证有其优点，但也并非毫无瑕疵。首先，u 中的众多因素可能各有极为不同的总体分布（比如在工资方程中，误差中的个人能力和教学质量）。但中心极限定理（CLT）在这些情形下仍成立，这种正态近似可能不那么好，这依赖于 u 中有多少因素以及它们的分布有多么地不同。

CLT 论证中更严重的问题是，它假定所有不可观测因素都以各自的和可加的方式影响着 y。对于这一点没有任何保证。如果 u 是不可观测因素的一个复杂函数，那么 CLT 论证并不真正适用。

在任何一个应用中，是否可以假定 u 的正态性，实际上都是一个经验性问题。例如，没有一个定理会认为取决于 *educ*，*exper* 和 *tenure* 的 *wage*

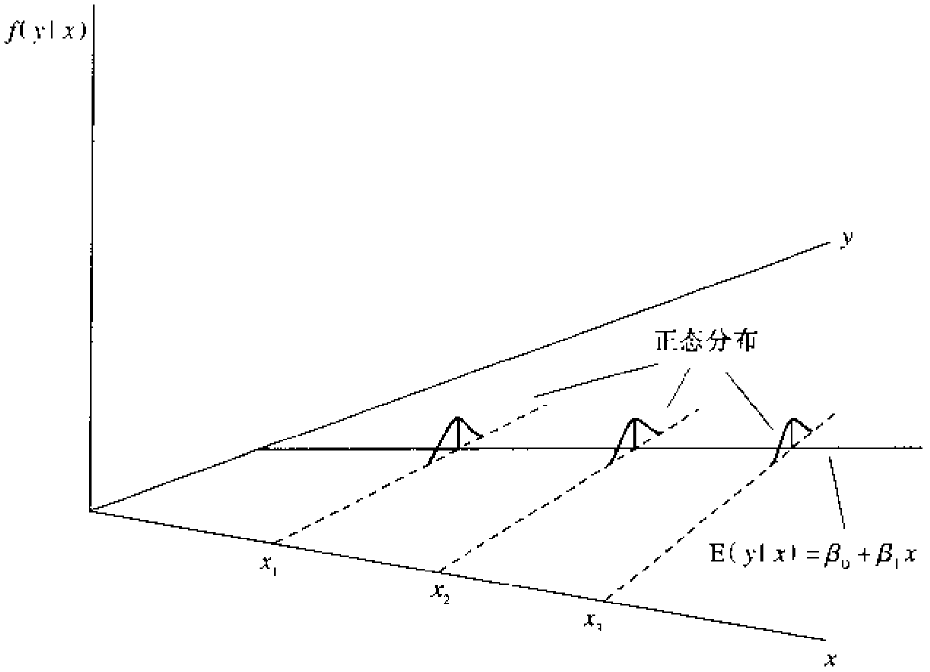


图4.1 单独一个解释变量的同方差正态分布

服从正态分布。如果有的话，那就是简单的推理表明相反的方面是成立的：由于工资绝不可能低于零，所以严格地讲，它不可能具有正态分布。而且，因为存在最低工资法，总体中有一定比例的人恰好得到最低工资，所以也与正态性假定相违背。尽管如此，作为一个实践问题，我们还是可以问条件工资分布是否“接近”正态分布。以往的经验证据表明，正态性假定对工资而言不是一个很好的假定。

通常利用一种变换（特别是取对数）就能得到一个更接近于正态的分布。比如，诸如log（*price*）之类的变量，常具有一个比*price*的分布更接近 115 正态的分布。同样，这又是一个经验问题，我们在第5章将进一步讨论。

有一些MLR.6明显不对的例子。要是y仅取少数几个值，它就不可能接近正态分布。例3.5中的因变量是一个很好的例子。一个年轻人在1986年被逮捕的次数这个变量*narr*86，取值仅限于一个很小的整数范围，而且对多数人来说都是零。因此，*narr*86远非正态分布。在这些情况下，该怎么做呢？如我们在第5章将看到的那样，相对于很大的样本容量来说，误差的非正态性算不上一个严重的问题，而且这一点很重要。目前来看，我们姑且认可正态性的假定。

误差项的正态性导致OLS估计量的正态抽样分布。

定理4.1（正态抽样分布）

在CLM假定MLR.1～MLR.6下，给定自变量的样本值，有

$$\hat{\beta}_j \sim \text{Normal}[\beta_j, \text{Var}(\hat{\beta}_j)] \tag{4.1}$$

116 式中，$\text{Var}(\hat{\beta}_j)$在第3章［方程（3.51）］中给出。因此

$$(\hat{\beta}_j-\beta_j)/\text{sd}(\hat{\beta}_j) \sim \text{Normal}(0,1)$$

给定附录B中正态分布随机变量的性质，证明定理（4.1）并不困难。每个$\hat{\beta}_j$都可以写成$\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij}u_i$。式中，$w_{ij} = \hat{r}_{ij}/\mathrm{SSR}_j$；$\hat{r}_{ij}$为$x_j$对所有其他自变量进行回归的第$i$个残差；而$\mathrm{SSR}_j$是这个回归的残差平方和［参见方程（3.62）］。因为w_{ij}只取决于自变量，所以它们可作为非随机变量来处理。因此，$\hat{\beta}_j$只是样本误差$\{u_i: i = 1, 2, \cdots, n\}$的一个线性组合。在假定MLR.6（和随机抽样假定MLR.2）下，误差是独立同分布的正态$(0, \sigma^2)$随机变量。有关独立正态随机变量的一个重要特征是，这种随机变量的线性组合仍是正态分布（参见附录B）。这就基本上完成了证明。在3.3节，我们证明了$\mathrm{E}(\hat{\beta}_j) = \beta_j$，而在3.4节，我们又推导了$\mathrm{Var}(\hat{\beta}_j)$；这里不需要重新推导这些结论。

问题 4.1

假设独立于解释变量，而与以相同的概率1/5取值-2，-1，0，1和2。这样会违背高斯-马尔科夫假定吗？会违背CLM假定吗？

如果通过将一个正态随机变量除以它的标准差而使之标准化，就得到一个标准正态随机变量，由此便直接得到定理的第二部分。

定理（4.1）的结论可以加强。除式（4.1）外，$\hat{\beta}_0$，$\hat{\beta}_1$，…，$\hat{\beta}_k$的任何线性组合也都是正态分布的，而且$\hat{\beta}_j$的任何一个子集也都有一个**联合**正态分布。这些结论就构成了本章剩余部分检验结论的基础。我们在第5章将证明，即使没有误差的正态性，在大样本情况下OLS估计量的正态性也会近似成立。

4.2 检验对单个总体参数的假设：t检验

本节包括了一个十分重要的专题，即对总体回归函数中有关某单个参数的假设进行检验。总体模型可写作

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{4.2}$$

而且假设它满足CLM假定。我们知道，OLS得到β_j的无偏估计量。在本节中，我们研究如何检验那些有关某个特定β_j的假设。为了充分了解假设检验，就必须记得，β_j是总体的未知特征，而且我们将永远不会确定地知道它们。尽管如此，我们还是可以对β_j的值作出假设，然后通过统计推断来检验我们的假设。

为了构造假设检验，我们需要如下结论。

定理 4.2（标准化估计量的t分布）

117 在CLM假定MLR.1～MLR.6下，有

$$(\hat{\beta}_j - \beta_j)/\mathrm{se}(\hat{\beta}_j) \sim t_{n-k-1} \tag{4.3}$$

式中，$k+1$为总体模型$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$中未知参数的个数（$k$个斜率参数和截距$\beta_0$）。

此定理在某些重要方面与定理 4.1 不同。定理 4.1 表明，在 CLM 假定下，$(\hat{\beta}_j-\beta_j)/\mathrm{sd}(\hat{\beta}_j)\sim \mathrm{Normal}(0,1)$。式(4.3)中的 t 分布源于如下事实，即 $\mathrm{sd}(\hat{\beta}_j)$ 中的常数 σ 已经被随机变量 $\hat{\sigma}$ 所取代。证明由此导致一个自由度为 $n-k-1$ 的 t 分布，并没有加深多少我们的见识。本质上讲，对它的证明表明，式（4.3）可写成标准正态随机变量 $(\hat{\beta}_j-\beta_j)/\mathrm{sd}(\hat{\beta}_j)$ 与 $\hat{\sigma}^2/\sigma^2$ 的平方根之比。可以证明二者是独立的，而且 $(n-k-1)\hat{\sigma}^2/\sigma^2\sim\chi^2_{n-k-1}$。于是根据 t 随机变量的定义，便得到这个结论（参见第 B.5 节）。

定理 4.2 的重要性在于，它使我们能检验有关 β_j 的假设。在多数应用中，我们主要的兴趣在于检验**虚拟假设**（null hypothesis）

$$\mathrm{H}_0:\ \beta_j=0 \tag{4.4}$$

式中，j 为对应着 k 个自变量中的任何一个。重要的是，要理解式（4.4）的含义，并能在一个特定的应用中用简单的语言来描述这个假设。由于 β_j 在控制了所有其他自变量后，度量了 x_j 对 y（的期望值）的偏效应，所以式（4.4）意味着，一旦对 $x_1, x_2, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k$ 都作出了解释，则 x_j 对 y 的期望值就没有任何偏效应。我们不能把虚拟假设表述成"x_j 对 y 有偏效应"，因为它对 β_j 不为零的任何一个值都成立。经典检验适合于检验像式（4.4）那样的简单假设。

作为一个例子，考虑工资方程

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 tenure+u$$

虚拟假设 H_0：$\beta_2=0$ 意味着只要对教育程度和现职任期进行了解释，工作年数（*exper*）对小时工资就没有影响。这是一个有经济意义的假设。如果它是正确的，那就意味着一个人在现任职之前的工作经历并不会影响工资。如果 $\beta_2>0$，则以前的工作经历会提高生产力，并因此提高工资。

你可能记得，在统计学教程中，学过对正态总体的均值进行假设检验的入门知识。（附录 C 复习了这部分内容。）在多元回归背景下检验式（4.4）的过程与此十分类似。虽然困难的部分在于得到系数估计值、标准误和临界值，但多数工作都可以由计量软件自动完成。我们的任务是，了解如何用回归结果来检验我们关心的假设。

我们用来检验式（4.4）（相对任何一个对立假设）的统计量被称为 β_j 的"所谓" **t 统计量**（t statistic）或"所谓" **t 比率**（t ratio），并被定义为

118

$$t_{\hat{\beta}_j}\equiv\frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j)} \tag{4.5}$$

我们把"所谓"字放在引号内是因为，如我们很快将看到的那样，要检验 β_j 的其他假设，还需要 t 统计量的一个更一般的形式。目前需要知道的是，式（4.5）适合于检验式（4.4）。在不引起混淆的情况下，我们有时将用 $t_{\hat{\beta}_j}$ 取代。

给定 $\hat{\beta}_j$ 及其标准误，$\hat{\beta}_j$ 的 t 统计量很容易计算。实际上，多数回归软件包都可以帮你计算，对每个系数都报告 t 统计量及其标准误。

在讨论如何用式（4.5）规范地检验 H_0：$\beta_j=0$ 之前，看看为什么 $t_{\hat{\beta}_j}$ 所

具有的性质能使之合理地成为一个侦查 $\beta_j \neq 0$ 的检验统计量，这会很有帮助。首先，因为 sd（$\hat{\beta}_j$）总为正，所以 $t_{\hat{\beta}_j}$ 与 $\hat{\beta}_j$ 的符号相同：如果 $\hat{\beta}_j$ 为正，则 $t_{\hat{\beta}_j}$ 也为正；如果 $\hat{\beta}_j$ 为负，则 $t_{\hat{\beta}_j}$ 也为负。其次，对于一个给定的 sd（$\hat{\beta}_j$）值，$\hat{\beta}_j$ 的值越大，则 $t_{\hat{\beta}_j}$ 的值也越大。如果 $\hat{\beta}_j$ 负得越多，则 $t_{\hat{\beta}_j}$ 也负得越多。

因为我们要检验的是 H_0：$\beta_j=0$，所以惟一自然的就是用 β_j 的无偏估计量 $\hat{\beta}_j$ 作为指导。在任何一个有意义的应用中，无论 H_0 是否正确，点估计值 $\hat{\beta}_j$ 都不可能正好等于零。问题是 $\hat{\beta}_j$ 与零相差有多大？$\hat{\beta}_j$ 的样本值与零相差很远就为拒绝 H_0：$\beta_j=0$ 提供了证据。但必须承认，在估计值 $\hat{\beta}_j$ 中也存在抽样误差，所以 $\hat{\beta}_j$ 的大小必须由其抽样误差来衡量。由于 $\hat{\beta}_j$ 的标准误是 $\hat{\beta}_j$ 标准差的一个估计值，所以 $t_{\hat{\beta}_j}$ 度量了被估计的标准差 $\hat{\beta}_j$ 与零相差多大。这恰好是我们利用初级统计学中标准的 t 统计量检验总体均值是否为零时所做的工作。$t_{\hat{\beta}_j}$ 的值充分远离零将导致拒绝 H_0。确切的拒绝法则取决于对立假设和所选择的检验的显著性水平。

在给定显著性水平（即当 H_0 正确时拒绝它的概率）下决定一个拒绝式（4.4）的法则，需要知道在 H_0 正确时 $t_{\hat{\beta}_j}$ 的抽样分布。我们从定理 4.2 知道，这就是 t_{n-k-1}。这正是检验式（4.4）所需要的主要理论结果。

在继续读下去之前，需要记住，我们正在检验的假设是关于**总体**参数的。我们**不是**在检验一个来自特定样本的估计值。因此，将一个虚拟假设表述成"H_0：$\hat{\beta}_j=0$"，或者在样本中一个参数的估计值是 0.237 时更糟糕地说"H_0：0.237＝0"，都是毫无意义的。我们要检验的是未知总体值 β_j 是否为零。

某些关于回归分析的讨论，将 t 统计量定义为式（4.5）的绝对值，从而 t 统计量总为正。这种做法的缺陷是，它使得对单侧对立假设检验十分累赘。在整章中，t 统计量总具有与对应 OLS 系数估计值相同的符号。

对单侧对立假设的检验

为了决定一个拒绝 H_0 的法则，我们需要决定相关的**对立假设**（alternative hypothesis）。首先考虑如下形式的一个**单侧对立假设**（one-sided alternative）

$$H_1:\ \beta_j>0 \tag{4.6}$$

119 这意味着，我们并不关心 H_0 形如 H_1：$\beta_j<0$ 的对立假设；出于某种原因，可能基于反思或经济理论，我们排除了 β_j 的总体值小于零的可能性。（对此的另一种看法是，虚拟假设实际上是 H_0：$\hat{\beta}_j\leqslant 0$；但无论在哪一种情况下，$t_{\hat{\beta}_j}$ 都被用做检验统计量。）

我们应该如何选择一个拒绝法则呢？我们必须首先决定一个**显著性水平**（significance level）或当 H_0 实际上正确时拒绝它的概率。为简洁起见，假设我们已决定了一个 5% 的显著性水平，这是最受欢迎的选择。因此，我们希望在 H_0 正确时只有 5% 的次数被错误地拒绝。在 H_0 下，$t_{\hat{\beta}_j}$ 有一个 t 分布

（因而均值为零），而在对立假设 $\beta_j>0$ 下，$t_{\hat{\beta}_j}$ 的期望值为正。因此，我们在寻找 $t_{\hat{\beta}_j}$ 的一个“足够大”的正值，以拒绝 H_0：$\beta_j=0$ 而支持 H_1：$\beta_j>0$

在5%的显著性水平上“足够大”的定义是，在含有 $n-k-1$ 个自由度的 t 分布中，处在百分位中第95位的数值，且用 c 表示。换句话说，**拒绝法则**（rejection rule）就是，在

$$t_{\hat{\beta}_j}>c \tag{4.7}$$

时，H_0 在5%的显著性水平上被拒绝并支持 H_1。通过我们对**临界值**（critical value）c 的选择，当 H_0 正确时，对所有随机样本有5%的可能会拒绝 H_0。

120 式（4.7）中的拒绝法则是**单侧检验**（one-tailed test）的一个例子。要得到 c，我们只需要显著性水平和自由度。例如，对于5%的显著性水平上的检验和 $n-k-1=28$ 个自由度，临界值是 $c=1.701$。如果 $t_{\hat{\beta}_j}<1.701$，那我们就不能在5%的显著性水平上拒绝 H_0 而支持式（4.6）。注意，当 $t_{\hat{\beta}_j}$ 为负时，无论其绝对值有多大，都不能拒绝 H_0 并支持式（4.6）。见图4.2。

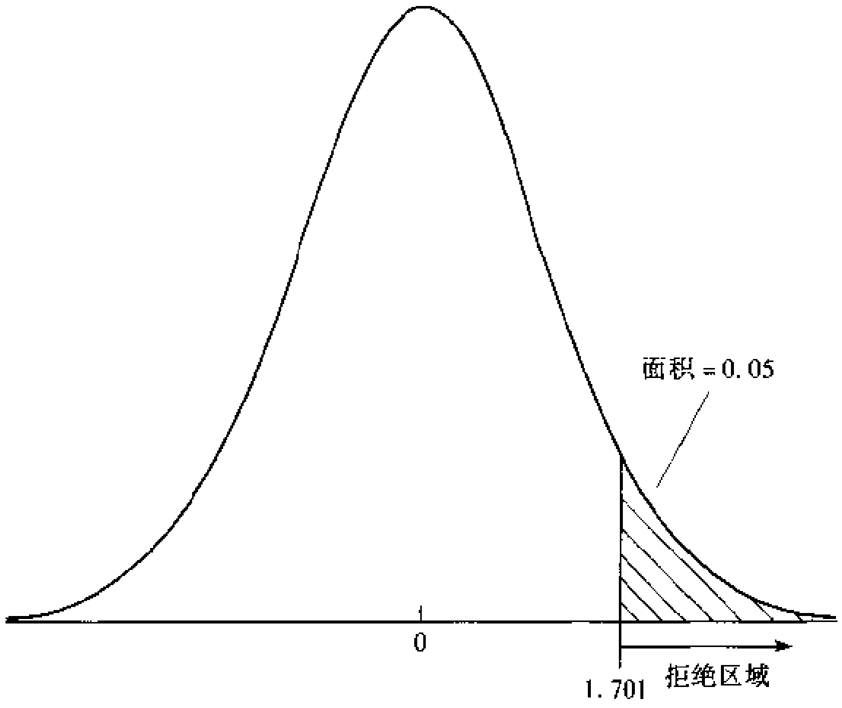


图4.2 在有28个自由度时拒绝对立假设 H_1：$\beta_j>0$ 的法则

同样的程序可用于其他的显著性水平。对一个显著性水平为10%的检验和自由度 $df=21$，其临界值为 $c=1.323$。对显著性水平为1%和 $df=21$，则 $c=2.518$。所有这些临界值都可以直接从表G.2得到。你应该注意到临界值中的一种模式：随着显著性水平下降，临界值会提高，以至我们要拒绝 H_0 就需要越来越大的 $t_{\hat{\beta}_j}$。因此，如果 H_0 在5%的显著性水平上被拒绝，那么它在10%的显著性水平上同样会自动被拒绝。在5%的显著性水平上拒绝了虚拟假设后，再重新去检验显著性水平为10%时的结果就毫无意义。

随着 t 分布的自由度逐渐变大，t 分布会接近标准正态分布。例如，当

$n-k-1=120$ 而显著性水平为 5%时，单侧对立假设（4.7）的临界值就是 1.658，标准正态的临界值则是 1.645。从实用的角度来看，它们相当接近；只要自由度大于 120，就可以使用标准正态的临界值。

例 4.1　小时工资方程

使用 WAGE1.RAW 中的数据得到如下估计方程：

$$\log(\hat{wage}) = 0.284 + 0.092educ + 0.0041exper + 0.022tenure$$
$$(0.104)(0.007)\qquad(0.0017)\qquad(0.003)$$
$$n = 526, R^2 = 0.316$$

标准误标在被估计参数下面的括号中，我们以后在全书中都采用这一做法。这个方程可用来检验在控制了 *educ* 和 *tenure* 的情况下，总体中 *exper* 的回报是否为零，以对立于回报为正的假设。可以写成 H_0：$\beta_{exper}=0$ 对 H_1：$\hat{\beta}_{exper}>0$。（在应用中，用相应变量名称做参数的下标是一个很好的方法，因为我们在一般模型中用做下标的数符都很随意，而且容易导致混淆。）记住，β_{exper}表示的未知总体参数。写“H_0：0.0041 = 0”或“H_0：$\hat{\beta}_{exper}=0$”是毫无意义的。

因为有 522 个自由度，所以可以使用标准正态的临界值。显著性水平为 5%的临界值是 1.645，显著性水平为 1%的临界值是 2.326。$\hat{\beta}_{exper}$ 的 t 统计量是

$$t_{\hat{\beta}_{exper}} = 0.0041/0.0017 \approx 2.41$$

所以 $\hat{\beta}_{exper}$ 或 *exper* 即使在 1%的显著性水平上都是统计显著的。我们也说“$\hat{\beta}_{exper}$在 1%的显著性水平上是统计显著大于零的”。

保持现职任期和受教育水平不变，多一年工作经历的估计回报并不大。比如，增加 3 年工作经历使 log（*wage*）提高 3（0.0041）= 0.0123，所以工资只高出 1.2%。不过，我们还是有说服力地证明了，总体中工作经历的偏效应是正的。121

应用中也会出现参数小于零的单侧对立假设：

$$H_1:\ \beta_j<0 \tag{4.8}$$

对立假设（4.8）的拒绝法则正好是前一种情形的映像。现在，临界值来自 t 分布的左侧。实践中，最简单的办法是将拒绝法则看成

$$t_{\hat{\beta}_j}<-c \tag{4.9}$$

式中，c 是对立假设 H_1：$\beta_j>0$ 的临界值。因为 t 分布表中只报告正的临界值，所以为简单起见，我们总是假定 c 为正，这样，临界值 $-c$ 也就是一个负数。

问题 4.2☞　令社区贷款许可率由下式决定：

$$apprate = \beta_0 + \beta_1 permin + \beta_2 avginc + \beta_3 avgwlth + \beta_4 avgdebt + u$$

式中，*permin* 为社区中少数民族所占的百分比；*avginc* 为平均收入；*avgwlth* 为平均财富；*avgdebt* 为对平均债务负担的某种度量。你如何在控制了平均收入、平均财富和平均债务的情况下，表述贷款率在各种族和民族构成间无差异的虚拟假设？你如何表述在贷款许可率上对少数民族存有歧视的对立假设？

比如，如果显著性水平是 5%，自由度为 18，那么 $c = 1.734$，所以如果$t_{\hat{\beta}_j} < -1.734$，在 5% 的显著性水平上就会拒绝 H_0：$\beta_j = 0$ 而支持 H_1：$\beta_j < 0$。重要的是记得，要用负的对立假设（4.8）来拒绝 H_0，我们必须得到一个负的统计量。正的 t 比率（无论多大）都不能为此提供证据。图 4.3 给出了其拒绝法则。

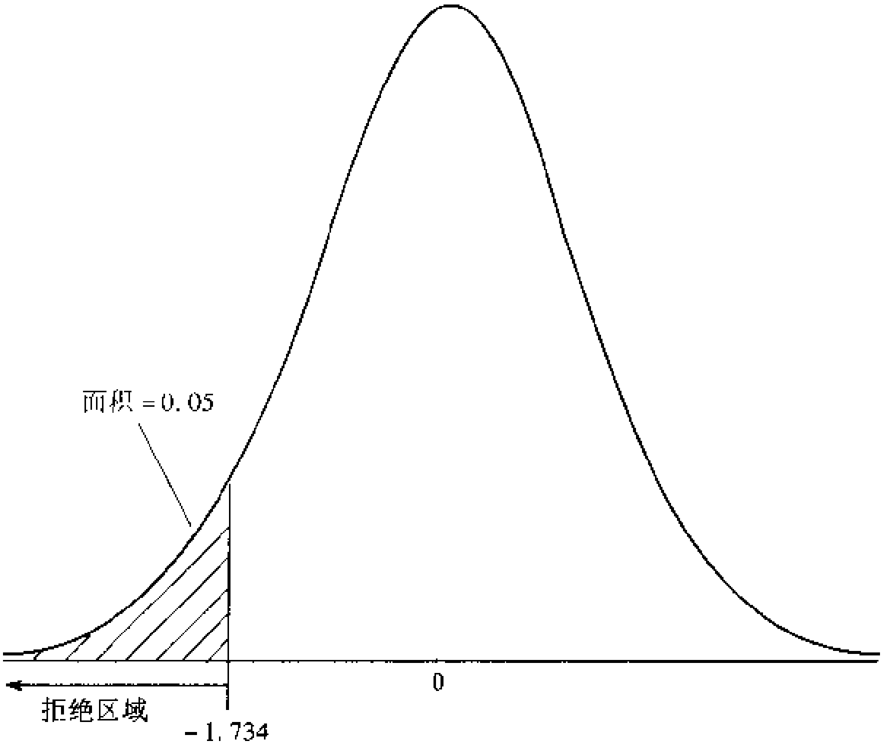


图 4.3　自由度为 18 和显著性水平为 5%，对立假设 H_1：$\beta_j < 0$ 的拒绝法则

例 4.2　学生表现与学校规模

学校规模对学生表现的影响很有意思。（参见《纽约时报》1995 年 5 月 28 日。）一种观点认为，在所有其他条件相同的情况下，小学校的学生比大学校的学生情况要好一些。即使在学校之间班级数量调整之后，仍认为这个假设是正确的。

文件 MEAP93.RAW 包含了 1993 年密歇根州 408 所高中的数据。我们可以用这些数据来检验虚拟假设：学校规模对标准化考试分数没有影响，对立假设是学校规模具有负效应。学生表现由学生通过密歇根教学评价委员会（MEAP）标准化十分制数学测验（*math*10）的百分比度量。学校规模由学

生注册人数（$enroll$）来度量。虚拟假设是 H_0：$\beta_{enroll}=0$，而对立假设是 H_1：$\beta_{enroll}<0$。目前，我们将控制两个其他因素：年均教师工资（$totcomp$）和平均每千名学生拥有的教职工人数（$staff$）。前者是对教师质量的一种度量，后者则大致度量了学生所得到的关注程度。

122 估计方程（标准误在括号中）是

$$\widehat{math10}=2.274+0.000\,46totcomp+0.048staff-0.000\,20enroll$$

$$(6.113)(0.000\,10)\qquad(0.040)\qquad(0.000\,22)$$

$$n=408,\ R^2=0.054\,1$$

$enroll$ 的系数 $-0.000\,2$ 与大学校有害于学生表现的猜测相一致：注册学生越多，学生通过十分制数学测验的百分比就越低。（$totcomp$ 与 $staff$ 的系数也具有我们预期的符号。）$enroll$ 的估计系数不同于零，可能仅仅因为抽样误差；为了使一种影响能有说服力，我们需要进行 t 检验。

由于 $n-k-1=408-4=404$，所以我们可以使用标准正态的临界值。在5%的显著性水平上，临界值是 -1.65；$enroll$ 的 t 统计量必须小于 -1.65，才能在5%的显著性水平上拒绝 H_0。

$enroll$ 的 t 统计量是 $-0.000\,2/0.000\,22\approx-0.91$，大于 -1.65：我们不能在5%的显著性水平上拒绝 H_0 并支持 H_1。实际上，显著性水平为15%时的临界值为 -1.04，而由于 $-0.91>-1.04$，所以我们即使在15%的显著性水平上也不能拒绝。我们断定，$enroll$ 在15%的显著性水平上都不是统计显著的。

由于变量 $totcomp$ 的 t 统计量是4.6，所以它即使在显著性水平为1%时
123 也是统计显著的。另一方面，变量 $staff$ 的 t 统计量是1.2，所以我们即使在显著性水平为10%时，我们也不能拒绝 H_0：$\beta_{staff}=0$ 而支持对立假设 H_1：$\beta_{staff}>0$。（标准正态分布的临界值是 $c=1.28$。）

为了解释函数形式如何影响我们的结论，我们将模型中的自变量都取对数后再进行估计。这样一来，比如说，就能使得随着学校规模的扩大，学校规模效应递减。估计方程为

$$\widehat{math10}=-207.66+21.16\log(totcomp)+3.98\log(staff)-1.29\log(enroll)$$

$$(48.70)\quad(4.06)\qquad(4.19)\qquad(0.69)$$

$$n=408,R^2=0.065\,4$$

log（$enroll$）的 t 统计量约是 -1.87；因为它低于显著性水平为5%时的临界值 -1.65，所以我们在5%的显著性水平上拒绝 H_0：$\beta_{\log(enroll)}=0$ 而支持对立假设 H_1：$\beta_{\log(enroll)}<0$。

在第2章，我们曾遇到一个因变量以原始形式（水平值形式）出现而自变量以对数形式出现的模型（被称为水平—对数模型）。在多元回归背景下，对参数的解释是一样的，当然不同之处仅在于，我们还能给出其他条件不变的解释。保持 $totcomp$ 和 $staff$ 不变，有 $\Delta\widehat{math10}=-1.29[\Delta\log(enroll)]$，

所以

$$\Delta\widehat{math}10 \approx \frac{1.29}{100\ (\%\Delta enroll)} \approx -0.013\ (\%enroll)$$

我们再次用到 log（*enroll*）的变化乘以 100 近似等于 *enroll* 的百分比变化这个近似公式。因此，如果一个学校的注册人数增加 10%，则预计 $\widehat{math}10$ 会下降 1.3%（*math*10 以百分数度量）。

我们会更喜欢哪个模型呢？使用 *enroll* 的水平值的模型，还是使用 log（*enroll*）的模型？在水平—水平模型中，注册人数并没有统计显著的影响，而在水平—对数模型中则有影响。这种函数形式的变化，导致水平—对数模型具有更高的 R^2，这意味着我们用 *enroll* 的形式能更多地解释 *math*10 的变异（从 5.4% 提高到 6.5%）。因为水平—对数模型更接近地刻画了 *math*10 与 *enroll* 之间的关系，所以颇受青睐。我们在第 6 章将更多地谈到如何使用 R^2 选择函数形式的问题。

双侧对立假设

在应用中，常常针对**双侧对立假设**（two-sided alternative）来检验虚拟假设 H_0：$\beta_j=0$，即

$$H_1: \beta_j \neq 0 \tag{4.10}$$

在这个对立假设下，x_j 对 y 具有未明确说明是正是负的影响。当经济理论（或常识）没有很好地说明 β_j 的符号时，这是一个恰当的对立假设。即便我们知道 β_j 在对立假设中是正或负，采取双侧检验通常也是明智的。最起码，使用双侧对立假设，能使我们避免检查估计方程并根据 $\hat{\beta}_j$ 的符号而提出对立假设。经典的统计推断要求我们在看到数据之前，就先表述虚拟假设和对
124 立假设，因此，企图使用回归估计值来帮助我们表述虚拟假设或对立假设，那是不允许的。比如，我们不应该先估计联系数学成绩和注册人数的方程，看到注册人数的估计效应为负后，再决定恰当的对立假设是 H_1：$\beta_{enroll}<0$

若对立假设是双侧的，则我们关心的是 t 统计量的**绝对值**。针对式（4.10），拒绝 H_0：$\beta_j=0$ 的法则是

$$|t_{\hat{\beta}_j}| > c \tag{4.11}$$

式中，$|\cdot|$ 为绝对值；而 c 为适当选取的临界值。为了找到 c，我们需要确定一个显著性水平，比方说 5%。对一个**双尾检验**（two-tailed test）而言，选择的 c 要使 t 分布两端的面积各等于 2.5%。换句话说，c 就是含有 $n-k-1$ 个自由度的 t 分布中的第 97.5 个百分位。当 $n-k-1=25$ 时，双侧检验在显著性水平为 5% 时的临界值是 2.060。图 4.4 给出了这个分布的一个说明。

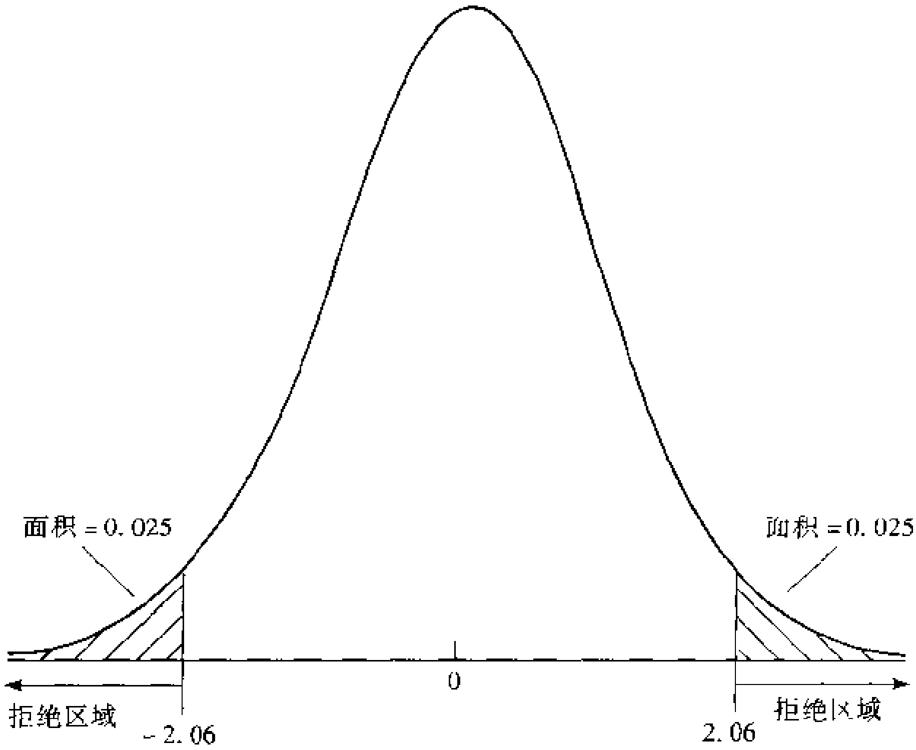


图 4.4　自由度为 25，显著性水平为 5%，对立假设为 H_1：$\beta_j \neq 0$ 时的拒绝法则

在没有明确地表述对立假设时，通常都认为是双侧的。在本书余下的内容中，没有给出对立假设，则表示是双侧对立假设；没有给出显著性水平，125 则表示显著性水平为 5%。在进行经验计量经济分析时，明确表述对立假设和显著性水平总是一个好主意。如果在 5% 的显著性水平上拒绝 H_0 而支持式（4.10），我们通常说"x_j 是**统计上显著的**（statistically significant），或在显著性水平为 5% 时统计上显著地异于零"。如果 H_0 未被拒绝，我们就说"x_j 在显著性水平为 5% 时是**统计上不显著的**（statistically insignificant）"。

例 4.3　大学 GPA 的决定因素

我们使用 GPA1.RAW 来估计一个解释大学 *GPA*（*colGPA*）的模型，并将平均每周缺课次数（*skipped*）作为一个新增的解释变量。估计模型是

$$\widehat{colGPA} = 1.39 + 0.412 hsGPA + 0.015 ACT - 0.083 skipped$$

$$(0.33)\ (0.094) \qquad (0.011) \qquad (0.026)$$

$$n = 141, R^2 = 0.234$$

通过对每个解释变量都使用双侧对立假设，能很容易地计算 t 统计量，以看出哪个变量是统计显著的。由于自由度很大（141 − 4 = 137），足以使用标准正态作为近似，所以显著性水平为 5% 的临界值约为 1.96，显著性水平为 1% 的临界值约为 2.58。

hsGPA 的 t 统计量是 4.38，在很小的显著性水平上都是显著的。于是，

我们说"*hsGPA* 在任何惯常的显著性水平上都是统计显著的"。*ACT* 的 t 统计量是 1.36，使用双侧对立假设，即使在显著性水平为 10%时，它也不是统计显著的。*ACT* 的系数实际上也很小：*ACT* 提高 10 分（这是一个很大的提高），预计 *colGPA* 也只会增加 0.15 分。因此，*ACT* 在实践中和统计上都是不显著的。

skipped 的系数的 t 统计量是 $-0.083/0.026=-3.19$，所以 *skipped* 在 1%的显著性水平上也是统计显著的（$3.19>2.58$）。这个系数意味着每周多旷一节课，预计会使 *colGPA* 降低约 0.083 分。于是，保持 *hsGPA* 和 *ACT* 不变，一个不旷课的学生与一个每周旷课 5 次的学生相比，预计在 *colGPA* 上会相差约 0.42 分。记住，这并非对某个学生而言，而是针对总体中学生的平均情况而言。

在这个例子中，对于每个变量，我们都能证明某个单侧对立假设是恰当的。使用双侧检验，变量 *hsGPA* 和 *skipped* 都十分显著，并具有我们预期的符号，所以，没有理由去做单侧检验。另一方面，针对一个单侧对立假设（$\beta_3>0$），*ACT* 在 10%的显著性水平上显著而在 5%的显著性水平上又不显著。但这并没改变 *ACT* 的系数相当小这一事实。

检验 $\boldsymbol{\beta}_j$ 的其他假设

尽管 H_0：$\beta_j=0$ 是最常见的假设，但我们有时也想检验 β_j 是否等于其他某个给定常数。两个常见的例子是 $\beta_j=1$ 和 $\beta_j=-1$。通常，如果虚拟假设表述为

126

$$H_0:\ \beta_j=a_j \tag{4.12}$$

式中，a_j 为我们假设 β_j 的值，那么相应的 t 统计量就是

$$t=\frac{\hat{\beta}_j-a_j}{\mathrm{se}(\hat{\beta}_j)}$$

和前面一样，t 度量的是 $\hat{\beta}_j$ 偏离 β_j 的假设值达到估计标准差的多少倍。一般的 t 统计量最好写成

$$t=\frac{\text{估计值}-\text{假设值}}{\text{标准误}} \tag{4.13}$$

在式（4.12）下，这个 t 统计量就是定理 4.2 中的 t_{n-k-1}。当 $a_j=0$ 时，就得到常用的 t 统计量。

我们可以用一般 t 统计量针对单侧或双侧对立假设做检验。比如，如果虚拟假设和对立假设分别是 H_0：$\beta_j=1$ 和 H_1：$\beta_j>1$，那我们得到单侧对立假设临界值的方法恰如从前一样：不同之处在于我们如何计算 t 统计量，而不在于如何得到适当的 c。如果 $t>c$，我们就拒绝 H_0 而支持 H_1。此时，我们将说"$\hat{\beta}_j$ 在适当的显著性水平上统计显著地大于 1"。

例 4.4　校园犯罪与注册人数

考虑联系大学校园内犯罪次数（*crime*）与学生注册人数（*enrollment*）的一个简单模型：

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u$$

这是一个常弹性模型，其中 β_1 是犯罪对注册人数的弹性。因为我们想像得到，随着学校规模的增大，总的犯罪次数预计也会随之增加，所以我们检验 H_0：$\beta_1 = 0$ 没有多大用处。要检验的一个更有意思的假设是，犯罪对注册人数的弹性是 1，H_0：$\beta_1 = 1$。这意味着，注册人数增加 1%，犯罪大致也增加 1%。一个值得注意的对立假设是 H_1：$\beta_1 > 1$，它意味着，注册人数增加 1%，会使校园犯罪的次数增加不止 1%。如果 $\beta_1 > 1$，那么，在相对的意义上（不仅在绝对的意义上），越大的校园，其犯罪就越成问题。看出这一点的一种方法是将上述方程取指数函数：

$$crime = \exp(\beta_0)\, enroll^{\beta_1} \exp(u)$$

（有关自然对数和指数函数的性质，可参见附录 A。）若 $\beta_0 = 0$ 且 $u = 0$，则这个方程的 $\beta_1 < 1$、$\beta_1 = 1$ 和 $\beta_1 > 1$ 三种情况都画在图 4.5 中。

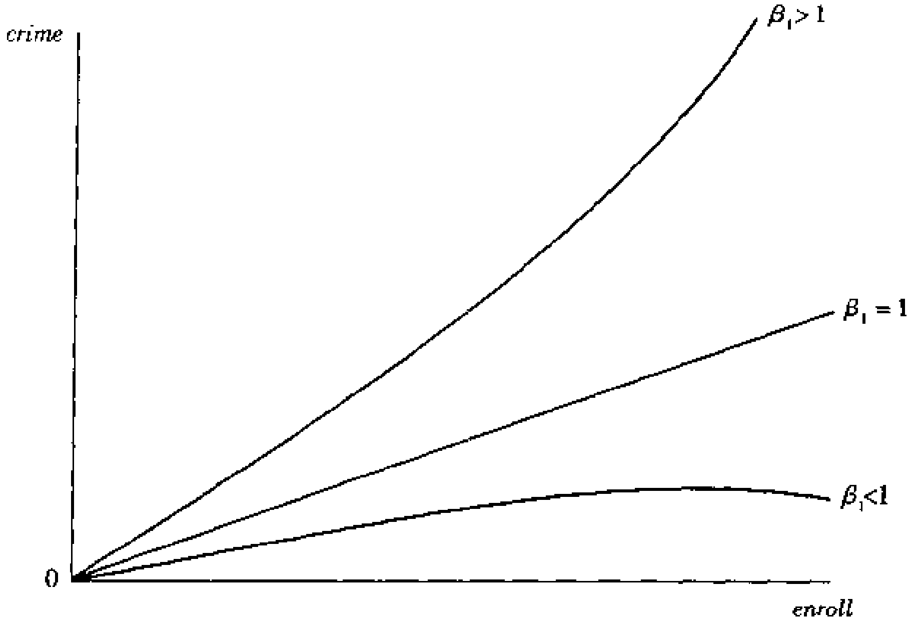


图 4.5　$crime = enroll^{\beta_1}$ 在 $\beta_1 < 1$，$\beta_1 = 1$ 和 $\beta_1 > 1$ 三种情况下的图示

我们利用美国 1992 年 97 个学院和大学的数据，针对 $\beta_1 > 1$ 来检验 $\beta_1 = 1$。数据来源于联邦调查局的《统一犯罪报告》（*FBI's Uniform Crime Reports*），样本中校园犯罪的平均次数约为 394 次，而平均注册人数约为 16 076人。估计方程（估计值和标准误都保留到小数点后两位）为

$$\widehat{\log(crime)} = -6.63 + 1.27\log(enroll)$$
$$(1.03) \quad (0.11) \qquad (4.14)$$

$$n = 97,\ R^2 = 0.585$$

犯罪对注册人数的估计弹性 1.27 处在对立假设 $\beta_1 > 1$ 的方向上。但是否有足够的证据断定 $\beta_1 > 1$ 呢？我们需要特别小心地检验这个假设，特别是因为，标准回归软件包给出的统计结果，比方程（4.14）所给出的简化结果要复杂得多。我们的第一反应可能是，用 log（*enroll*）的系数除以其标准误来构造“所需要的” t 统计量，这也是回归软件包所报告的 t 统计量。但这个统计量对于检验 H_0：$\beta_1 = 1$ 来说是错误的。正确的 t 统计量来自式（4.13）：从估计值中减去其假设值 1，然后除以 $\hat{\beta}_j$ 的标准误：$t = (1.27 - 1)/0.11 = 0.27/0.11 \approx 2.45$。含有 $97 - 2 = 95$ 个自由度的 t 分布，显著性水平为 5% 的单侧检验的临界值约为 1.66（使用 $df = 120$），所以我们在显著性水平为 5% 时明确拒绝 $\beta_1 = 1$ 而支持 $\beta_1 > 1$。事实上，1% 的临界值约为 2.37，所以即使在 1% 的显著性水平上也会拒绝虚拟假设而支持对立假设。

我们应该清楚，这一分析没有保持其他因素不变，所以弹性 1.27 不一定是在其他条件不变情况下很好的估计值。较大的注册人数可能与其他导致 128 较高犯罪的因素相关：学校越大，可能所处区域的犯罪率越高。我们可以通过搜集学校所在地的犯罪率数据来控制这一因素。

对于双侧对立假设，比如 H_0：$\beta_j = -1$ 和 H_1：$\beta_j \neq -1$，我们仍像在式（4.13）中一样计算 t 统计量：$t = (\hat{\beta}_j + 1)/\mathrm{se}(\hat{\beta}_j)$（注意减去 -1 就是加上 1）。拒绝法则就是双侧检验最常用的一个：若 $|t| > c$，其中 c 是双尾检验的临界值，则拒绝 H_0。如果被拒绝，我们就说，在适当的显著性水平上“$\hat{\beta}_j$ 统计显著异于 -1”。

例 4.5　住房价格和空气污染

对于一个由波士顿地区 506 个社区组成的样本，我们估计了一个联系社区中平均住房价格（*price*）与各种社区特征的模型：*nox* 表示空气中氧化亚氮的含量，以每区的百万分子数度量；*dist* 表示该社区相距五个商业中心的加权距离，以英里为单位；*rooms* 表示该社区平均每套住房的房间数；而 *stratio* 则为该社区学校的平均学生—教师比。总体模型为

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 \log(dist) + \beta_3 rooms + \beta_4 stratio + u$$

于是 β_1 是 *price* 对 *nox* 的弹性。我们希望针对对立假设 H_1：$\beta_1 \neq -1$ 来检验 H_0：$\beta_1 = -1$。做这个检验的 t 统计量是 $t = (\hat{\beta}_1 + 1)/\mathrm{se}(\hat{\beta}_1)$。

利用 HPRICE2.RAW 中的数据，估计模型为

$$\widehat{\log(price)} = 11.08 - 0.954\log(nox) - 0.134\log(dist)$$

$$(0.32)\quad (0.117)\qquad (0.043)$$

$$+\,0.255\,rooms - 0.052\,stratio$$

$$(0.019)\qquad (0.006)$$

$$n = 506,\ R^2 = 0.581$$

斜率估计值都具有预期的符号。每个系数都是在很小的显著性水平上统计显著地异于零，包括$\log(nox)$的系数。但我们不想检验$\beta_1 = 0$。我们所关心的虚拟假设是H_0：$\beta_1 = -1$，对应的t统计量是$(-0.954+1)/0.117 = 0.393$。当t统计量这么小时，几乎不需要看t分布中的临界值：即使在很大的显著性水平上，估计的弹性也不会统计显著地异于-1。控制了我们已经包括进来的因素，几乎没有证明这个弹性异于-1的证据。

计算 t 检验的 p 值

到目前为止，我们已经讨论了如何使用一个经典方法来进行假设检验：在陈述一个对立假设后，选择一个显著性水平，这个显著性水平又决定了一个临界值。一旦临界值确定了，将t统计量的值与这个临界值相比较，于是在给定的显著性水平下，虚拟假设要么被拒绝，要么未被拒绝。

129 即使在决定了适当的对立假设之后，经典方法中还有一种任意性成分，即我们必须提前选择一个显著性水平。不同的研究者根据特定的应用，会偏好不同的显著性水平。不存在一个“正确的”显著性水平。

事前指定一个显著性水平，可能隐藏假设检验结果方面的有用信息。比如，假使我们想针对一个双侧对立假设来检验一个参数为零的虚拟假设，自由度为40时我们得到t统计量等于1.85。由于这个t统计量小于双尾检验的临界值$c = 2.021$，所以这个假设在5%的显著性水平上未被拒绝。一个不准备拒绝虚拟假设的研究者可能仅仅需要与估计值一起报告这个结果即可：在5%的显著性水平上这个假设未被拒绝。当然，如果报告了t统计量（或系数及其标准误），那么，由于显著性水平为10%的临界值是1.684，所以我们也可以决定，虚拟假设在10%的显著性水平上被拒绝。

与其在不同的显著性水平上进行检验，不如回答如下更富于信息的问题：给定t统计量的观测值，能拒绝虚拟假设的最小显著性水平是多少？这个水平被称为检验的 **p 值**（p-value）（参见附录C）。在上一例中，由于虚拟假设在显著性水平为5%时未被拒绝，所以我们知道p值大于0.05；而由于虚拟假设在10%的水平上被拒绝，所以我们知道p值小于0.10。通过计算t随机变量（有40个df）在绝对值上大于1.85的概率，我们就能得到实际的p值。也就是说，p值就是当我们用检验统计量的值（上例中的1.85）作为检验临界值时的检验的显著性水平。这个p值示于图4.6。

因为p值是一个概率，所以它的值总是介于0～1之间。为了计算p值，我们要么需要极为详尽的t分布表（不是十分可行），要么需要一个计算程序来计算t分布的概率分布函数下的面积。许多回归软件包都具有后面这种能力。虽然有些软件包例行对每个OLS回归都计算p值，但也都是针对特定的假设检验。如果一个回归软件包与标准OLS结果一起报告了p值，

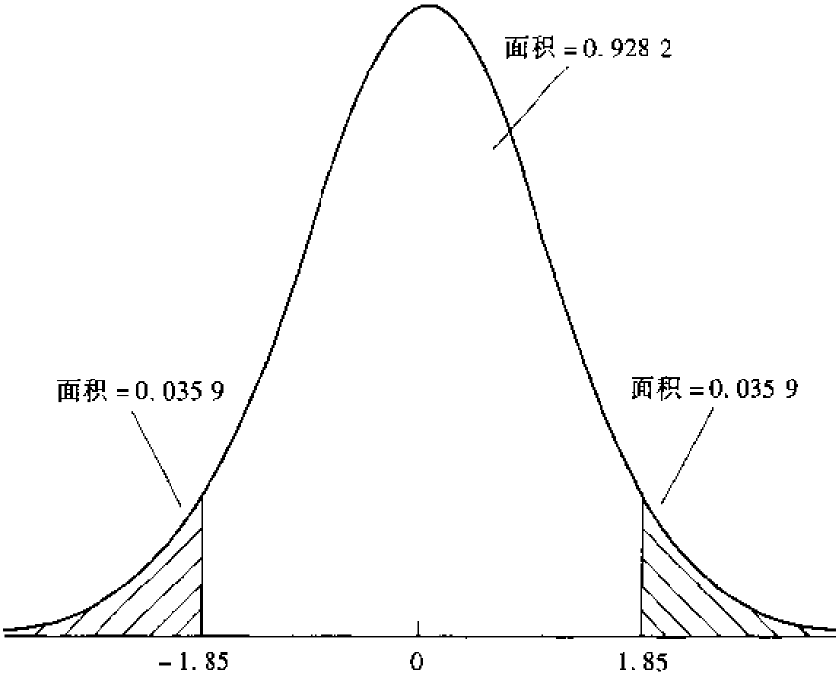


图 4.6　当 $t=1.85$ 和 $df=40$ 时，相对双侧对立假设所得到的 p 值

那么几乎可以肯定，这个 p 值是针对双侧对立假设来检验虚拟假设 H_0: $\beta_j=0$的。在这种情况下的 p 值是

$$P(|T|>|t|) \tag{4.15}$$

为清楚起见，我们令 T 表示一个自由度为 $n-k-1$ 的 t 分布随机变量，而令 t 表示该检验统计量的数值。

p 值很好地总结了经验证据拒绝虚拟假设的强弱。可能最有用的解释如下：p 值是我们观察到一个 t 统计量至少和在虚拟假设正确时的 t 统计量一样大的概率。这意味着，小 p 值是拒绝虚拟假设的证据；大 p 值不能提供拒绝 H_0 的证据。例如，如果 p 值 = 0.50（总是以小数报告，而不是以百分比报告），那么，我们将观察到一个 t 统计量的值，它至少和虚拟假设正确时在所有随机样本的 50% 中都能观察到的 t 统计量一样大；这是拒绝 H_0 的相当弱的证据。

130　在 $df=40$ 和 $t=1.85$ 的例子中，计算出来的 p 值是

$$p\text{ 值}=P(|T|>1.85)=2P(T>1.85)=2(0.035\,9)=0.071\,8$$

式中，$P(T>1.85)$ 为一个自由度为 40 的 t 分布中 1.85 以右部分的面积。（这个数值由计量经济软件 Stata 计算而来，表 G.2 中并没有这个值。）这意味着，如果虚拟假设正确，那么我们约有 7.2% 次观察到 t 统计量的绝对值至少和 1.85 一样大。虽然这为拒绝虚拟假设提供了一些证据，但在 5% 的显著性水平上，我们还不能拒绝这个虚拟假设。

前例说明了，一旦 p 值计算出来，在任何理想的显著性水平下都能进行经典检验。如果用 α 表示检验的显著性水平（以小数形式表示），那么，

若 p 值 $<\alpha$，则拒绝虚拟假设；否则，在 $100\alpha\%$ 的显著性水平下，就不能拒绝 H_0。

131 对单侧对立假设计算 p 值也相当简单。比如，假设对 H_1：$\beta_j>0$ 检验 H_0：$\beta_j=0$。如果 $\hat{\beta}_j<0$，计算 p 值就无关紧要：我们知道 p 值大于 0.50，这绝不会导致我们拒绝 H_0 而支持 H_1。如果 $\hat{\beta}_j>0$，那么 $t>0$，p 值恰好就是一个含有适当自由度的 t 随机变量超过 t 值的概率。一些回归软件包只对双侧对立假设计算 p 值。但要得到单侧对立假设的 p 值也很简单：只须将双侧对立假设的 p 值除以 2 即可。

问题 4.3 假设你估计了一个回归模型，得到 $\hat{\beta}_1=0.56$ 和针对 H_1：$\beta_1\neq 0$ 检验 H_0：$\beta_1=0$ 的 p 值 $=0.086$，那么针对 H_1：$\beta_1>0$ 检验 H_0：$\beta_1=0$ 的 p 值是多少？

如果对立假设是 H_1：$\beta_j<0$，在 $\hat{\beta}_j<0$（因此 $t<0$）时计算 p 值是有意义的：由于 t 分布关于零对称，所以 p 值 $=P(T<t)=P(T>|t|)$。同样，它也等于双尾检验 p 值的一半。

由于你很快就会熟悉导致统计显著性的 t 统计量的大小，特别是对于大样本容量，所以对 t 统计量报告 p 值并非至关重要，但报告了 p 值也没有害处。而且，当我们在 4.5 节讨论 F 检验时，我们将看到，由于 F 检验的临界值不那么容易记得，所以计算 p 值还是重要的。

对经典假设检验用语的提醒

当 H_0 未被拒绝时，我们喜欢说“在 $x\%$ 的水平上不能拒绝 H_0”，而不是说“我们在 $x\%$ 的水平上接受了 H_0”。我们用例 4.5 可以说明为什么前一个说法更好。在这个例子中，*price* 对 *nox* 的估计弹性是 -0.954，检验 $H_0:\beta_{nox}=-1$ 的 t 统计量是 0.393；因此，我们不能拒绝 H_0。但 β_{nox} 还有许多其他值（不计其数）都不能被拒绝。比如，$H_0:\beta_{nox}=-0.9$ 的 t 统计量是 $(-0.954+0.9)/0.117=0.462$，所以这个虚拟假设也不能被拒绝。显然 $\beta_{nox}=-1$ 和 $\beta_{nox}=-0.9$ 不可能同时正确，所以说我们“接受”这些假设中的每一个就毫无意义，所能说的只是数据使我们不能在 5% 的显著性水平上拒绝这些假设中的每一个。

经济或实际显著性与统计显著性

由于我们在整个一节都在强调统计显著性，除了 t 统计量的大小外，现在该是我们注意系数估计值大小的时候了。一个变量 x_j 的统计显著性完全由 $t_{\hat{\beta}_j}$ 的大小来决定，而一个变量的**经济显著性**（economic significance）或**实际显著性**（practical significance）则与 $\hat{\beta}_j$ 的大小（及符号）有关系。

记得检验 H_0：$\beta_j=0$ 时的 t 统计量被定义为估计值与其标准误之比：$t_{\hat{\beta}_j}=\hat{\beta}_j/\mathrm{se}(\hat{\beta}_j)$。之所以 $t_{\hat{\beta}_j}$ 能标志统计显著性，要么是因为 $\hat{\beta}_j$ “很大”，要么是因为 $\mathrm{se}(\hat{\beta}_j)$ “很小”。在实践中，区分导致 t 统计量统计显著的原因很重要。过多地强调统计显著性，即使一个变量的估计效应不太大，也认为它在解释 y 时很“重要”，从而导致错误的结论。

例 4.6　401k 养老金计划的参与率

在例 3.3 中，我们使用 401k 养老金方案中的数据估计了一个模型，用企业的贡献率和工人的年龄来描述养老金计划的参与率。我们现在再引进一个对企业规模的某种度量，即企业雇员总数（*totemp*）。估计方程变成

$$\widehat{prate}=80.29+5.44\,mrate+0.269\,age-0.000\,13\,totemp$$
$$(0.78)\quad(0.52)\qquad(0.045)\qquad(0.000\,04)$$
$$n=1\,534,\quad R^2=0.100$$

变量 *totemp* 的 t 统计量绝对值最小：$t=-0.000\,13/0.000\,04=-3.25$，而且在很小的显著性水平上都是统计显著的。（这个 t 统计量的双尾检验的 p 值约为 0.001。）因此，在相当小的显著性水平上，所有的变量都是统计显著的。

从实践的意义上看，*totemp* 的系数有多大呢？保持 *mrate* 和 *age* 不变，如果一个企业增加 10 000 个雇员，参与率也只下降 10 000（0.000 13）= 1.3 个百分点。雇员人数如此巨大的提高，对参与率只有很有限的影响。因此，企业规模确实会影响参与率，但这种影响在实践中并不是很大。

上例表明，在处理大样本时，除了看 t 统计量外，对系数的大小加以解释也特别重要。对于大样本容量，参数可以估计得相当准确：标准误与系数估计值相比通常都相当小，从而常常导致统计显著性。

一些研究者坚持在样本容量逐渐扩大时使用越来越小的显著性水平，部分原因就是要作为抵偿标准误越来越小的一种办法。比如，当 n 为几百时，我们感觉 5% 的显著性水平就满意了，而当 n 为几千时，可能要使用 1% 的显著性水平。使用较小的显著性水平意味着经济上和统计上的显著性更可能达成一致，尽管不能保证；在上例中，尽管使用的显著性水平小到 0.1%（1/10 个百分点），但仍能得到 *totemp* 统计上显著的结论。

大多数研究者还愿意在小样本容量的应用中接受较大的显著性水平，这反映了对较小样本容量更难发现显著性这一事实（临界值更大而估计量更不精确）。不幸的是，事实是否如此，还取决于研究者的基本计划安排。

例 4.7　在职培训津贴对企业废弃率的影响

一个制造企业的废弃率是每 100 个产品中因有缺陷而必须丢弃的产品个数。因此，废弃率的下降反映出更高的生产力。

我们用废弃率来度量对工人进行生产力培训的效果。对于 1987 年密歇

根州制造企业的一个样本，我们估计了如下方程：

$$\log(\hat{scrap}) = 13.72 \quad 0.028 hrsemp \quad 1.21\log(sales) + 1.48\log(employ)$$

$$(4.91)\ (0.019) \qquad (0.41) \qquad (0.43)$$

$$n - 30,\ R^2 = 0.431$$

（该回归使用了 JTRAIN.RAW 中数据的一个子集。）变量 *hrsemp* 为平均每个 133 雇员每年受到培训的小时数；*sales* 为企业年销售额；*employ* 为企业雇员人数。样本中的平均废弃率约为 3.5，而平均的 *hrsemp* 约为 7.3。

我们主要关心的变量是 *hrsemp*。每个雇员多受 1 个小时的培训，则使 log(*scrap*)下降 0.028，这意味着废弃率约低 2.8%。于是，若 *hrsemp* 增加 5 个单位——每个雇员每年多培训 5 个小时，则估计废弃率会下降 5(2.8) = 14%。虽然这看起来像是相当大的效果，但企业增加培训是否值得，却取决于培训的成本和降低废弃率的收益。尽管我们没有所需要的数据来进行成本收益分析，但估计的效果看来非同小可。

培训变量的**统计显著性**如何呢？*hrsemp* 的 t 统计量为 $-0.028/0.019 = -1.47$，现在你可能认识到，其大小不足以得到如下结论：*hrsemp* 在 5% 的显著性水平上是统计显著的。实际上，当单侧对立假设 $H_1: \beta_{hrsemp} < 0$ 的自由度为 $30 - 4 = 26$ 时，显著性水平为 5% 的临界值约为 -1.71。因此，使用一个显著性水平严格为 5% 的检验，即使使用单侧对立假设，我们也必然得到 *hrsemp* 在统计上不显著的结论。

因为样本容量相当小，所以我们可能对显著性水平的要求要宽松一些。显著性水平为 10% 的临界值是 -1.32，所以 *hrsemp* 在显著性水平为 10% 时，针对单侧对立假设是显著的。很容易计算出 p 值 $P(T_{26} < -1.47) = 0.077$。这个 p 值可能足够低，以致能断定培训的估计效果不仅仅是抽样误差所致。但有些经济学家对此存有异议。

我们知道，即使样本容量看上去相当大，很大的标准误仍可能是多重共线性（某些自变量之间高度相关）造成的结果。如我们在 3.4 节中谈到的那样，对于这个问题，我们除了搜集更多的数据或通过从模型中去掉一些自变量而改变分析范围外，就无能为力了。至于在小样本容量的情形中，当解释变量高度相关时，很难精确地估计其偏效应。（4.5 节中就有一个例子来说明这个问题。）

在本节结束之际，我们给出一些准则，以便讨论一个变量在多元回归模型中的经济和统计显著性。

1. 检查统计显著性。如果该变量是统计显著的，那就讨论系数的大小，以对其实际或经济上的重要性有所认识。下一步要特别小心，它取决于自变量和因变量出现在方程中的方式。（特别是，度量单位是什么？变量是不是以对数形式出现？）

2. 如果一个变量在通常的显著性水平（10%，5%或 1%）上不是统计显著的，那你仍可能要问，这个变量对 y 是否具有预期的影响，而这个影响在实践

中是否很大。如果这种影响很大，那你就应该对 t 统计量计算一个 p 值。对于小样本容量，你有时可以让 p 值大到 0.20（但没有一成不变的准则）。p 值大即 t 统计量小、实际上大的估计值可能来自抽样误差，所以我们如履薄冰：不同的随机样本可以导致极为不同的估计值。

134 3. 通常发现，t 统计量很小的变量都具有“错误”的符号。实际上，这些情况可以忽略：我们得出了这些变量在统计上不显著的结论。一个有出乎意料的符号而在实践中具有很大影响的显著变量，才成为问题而且难以解决。为了解决这种问题，人们通常要对模型和数据的性质做更多的思考。一个违背直觉而又显著的估计值，常常是因为遗漏了一个关键变量，或出现了我们在第 9 章至第 15 章将讨论的某个重要问题。

4.3 置信区间

在经典线性模型的假定之下，我们能很容易地为总体参数 β_j 构造一个**置信区间**（confidence interval，CI）。因为置信区间为总体参数的可能取值提供了一个范围，而不只是一个点估计值，所以又被称为**区间估计（值）**。

利用 $(\hat{\beta}_j-\beta_j)/\mathrm{se}(\hat{\beta}_j)$ 服从自由度为 $n-k-1$ 的 t 分布［参见式（4.3）］这个事实，经简单计算就能得到未知 β_j 的一个 CI。一个 95% 的置信区间就是

$$\hat{\beta}_j \pm c\cdot\mathrm{se}(\hat{\beta}_j) \tag{4.16}$$

式中，常数 c 是一个 t_{n-k-1} 分布的第 97.5 个百分位。更准确地说，置信区间的下界和上界分别是

$$\underline{\beta}_j \equiv \hat{\beta}_j - c\cdot\mathrm{se}(\hat{\beta}_j)$$

和

$$\bar{\beta}_j \equiv \hat{\beta}_j + c\cdot\mathrm{se}(\hat{\beta}_j)$$

到目前为止，回顾一下置信区间的含义很有好处。如果一次又一次地获得随机样本，每次都计算出 $\underline{\beta}_j$ 和 $\bar{\beta}_j$，那么，（未知的）总体值 β_j 将在 95% 的样本的区间（$\underline{\beta}_j$，$\bar{\beta}_j$）中出现。不幸的是，对于用来构造 CI 的单一个样本，我们并不知道 β_j 是否确实包含在这个区间中。我们希望得到的样本是所有样本中属于区间估计值包括了 β_j 的那 95% 中的一个，但我们没有把握。

利用当代计算技术构造一个置信区间是很简单的，总共需要三个量：$\hat{\beta}_j$，$\mathrm{sd}(\hat{\beta}_j)$ 和 c。任何一个回归软件包都会报告系数估计值及其标准误。为了得到，我们必须知道自由度 $n-k-1$ 和置信水平（这里是 95%）。然后，c 的值就可以从 t_{n-k-1} 分布中得到。

例如，若 $df=n-k-1=25$，则对任何一个 β_j，其 95% 的置信区间都

是$[\hat{\beta}_j - 2.06 \cdot \mathrm{se}(\hat{\beta}_j), \hat{\beta}_j + 2.06 \cdot \mathrm{se}(\hat{\beta}_j)]$。

当 $n-k-1>120$ 时，由于 t_{n-k-1}分布充分接近正态分布，所以我们可以使用标准正态分布中的第 97.5 个百分位，来构造一个置信水平为 95%的 CI：$\hat{\beta}_j \pm 1.96 \cdot \mathrm{se}(\hat{\beta}_j)$。实际上，当 $n-k-1>50$ 时，由于 c 的值与 2 如此接近，所以我们可以使用一个简单的**经验法则**，来构造置信水平为 95%的置信区间：$\hat{\beta}_j$ 加上或减去其 2 倍的标准误。对于自由度较小的情况，准确的百分位应该从 t 分布表中得到。 135

对任何其他的置信水平，构造置信区间也很容易。比如，通过在 t_{n-k-1}分布中选择第 95 个百分位作为 c，就可以得到一个置信水平为 90%的 CI。当 $df=n-k-1=25$ 时，$c=1.71$，所以置信水平为 90%的 CI 就是$\hat{\beta}_j \pm 1.71 \cdot \mathrm{se}(\hat{\beta}_j)$，它必须比置信水平为 95%的 CI 要窄些。对于置信水平为 99%的 CI，c 就是 t_{25}分布中的第 99.5 个百分位。当 $df=25$ 时，置信水平为 99%的 CI 大致是$\hat{\beta}_j \pm 2.79 \cdot \mathrm{se}(\hat{\beta}_j)$，毫无疑问，它比置信水平为 95%的 CI 宽些。

许多现代回归软件包都对每个系数及其标准误报告了一个置信水平为 95%的 CI，省得我们再做任何计算。一旦构造了一个置信区间，进行双尾假设检验就很容易。如果虚拟假设是 H_0：$\beta_j = a_j$，那么，当且仅当 a_j 不在置信水平为（比方说）95%的置信区间中时，相对于 H_1：$\beta_j \neq a_j$ 的 H_0 才会被拒绝。

例 4.8 住房的享受定价模型

一个用商品特性来解释其价格的模型被称为**享受定价模型**（hedonic price model）。如下方程就是住房价格的一个享受定价模型。住房特征包括面积的平方英尺数（*sqrft*）、卧室的间数（*bdrms*）和卫生间的个数（*bthrms*）。通常 *price* 和某些解释变量都以对数的形式出现。使用 1990 年在马萨诸塞州沃尔瑟姆销售的 19 套住房的观测数据，得到估计方程（标准误写在系数估计值下面的括号内）

$$\log(\widehat{price}) = \underset{(1.15)}{7.46} + \underset{(0.184)}{0.634}\log(sqrft) - \underset{(0.059)}{0.066}bdrms + \underset{(0.075)}{0.158}bthrms$$

$$n=19,\ R^2=0.806$$

由于 *price* 和 *sqrft* 都以对数的形式出现，所以价格对平方英尺数的弹性是 0.634，于是，保持卧室数和卫生间数不变，面积增加 1%预计会提高住房价格约 0.634%。注意到被估计模型的自由度为 $n-k-1=19-3-1=15$，我们便可以为总体弹性构造一个置信水平为 95%的置信区间。我们在表 G.2 中找到 t_{15}分布中的第 97.5 个百分位数：$c=2.131$。因此 $\beta_{\log(sqrft)}$置信水平为 95%的置信区间是 $0.634 \pm 2.131\ (0.184)$或$(0.242, 1.026)$。由于这个置信区间中不包括零，所以在 5%的显著性水平下通过双侧对立假设拒绝了 H_0：$\beta_{\log(sqrft)} = 0$。

bdrms 的系数为负看来有些违反直觉。但重要的是要记得，这个系数的

其他条件不变性：它度量的是在保持房间大小和卫生间个数不变的情况下，增加一间卧室的影响。如果两套住房面积一样大，而卧室的数量不同，那么卧室的间数较多的住房的卧室肯定面积较小；房间多一些而又小一些未必就是好事情。无论如何，我们可以看到，β_{bdrms}在置信水平为95%时的置信区间相当宽，而且还包含了零：0.066±2.131（0.059）或（−0.192，0.060）。因此，*bdrms* 在其他条件不变的情况下，对住房价格没有统计显著的影响。

136 给定住房面积和卧室间数，多一个卫生间预计会提高住房价格15.8%。（记住，我们必须将 *bthrms* 的系数乘以100才将其影响变成百分数。）β_{bthrms}在置信水平为95%时的置信区间是（−0.002，0.318）。在这种情况下，零刚刚处在置信区间中，所以从技术上讲，$\hat{\beta}_{bthrms}$相对双侧对立假设在5%的显著性水平上不算统计显著的。因为它极其接近于显著，所以我们可能会得出结论，卫生间的个数对log（*price*）有影响。

你应该记得，置信区间只能像用以构造它的基本假定那样好。如果我们漏掉了与解释变量相关的重要因素，那么系数估计值就不可靠：OLS是有偏的。如果出现了异方差性［比方说在前例中，如果log(*price*)的方差取决于任何一个解释变量］，那么标准误就不能作为$\mathrm{sd}(\hat{\beta}_j)$的一个估计值（如我们在3.4节所讨论的那样），使用这些标准误计算而得来的这个置信区间，也就不是一个置信水平为95%的CI。尽管我们在得到这些CI时也已经用到了对误差的正态性假定，但如我们在第5章将见到的那样，对于一个涉及几百个观测的应用案例来说，这一假定并不是那么重要。

4.4 检验关于参数的一个线性组合的假设

前两节已经表明，如何使用经典假设检验或置信区间来检验对单个参数β_j的假设。在应用中，我们常常需要检验涉及不止一个总体参数的假设。在本节中，我们要说明的问题是，如何对涉及不止一个参数β_j的单个假设进行检验。4.5节则说明，如何对多个假设进行检验。

为了说明这个一般方法，我们将考虑一个简单模型，以比较两年制大专和四年制本科教育的回报；为简明起见，我们把后者称为“大学”教育。［此例由凯恩和劳斯（Kane and Rouse，1995）提出，他们对这个问题进行了详尽的分析。］总体中包括了具有高中学历的工人，这个模型就是

$$\log(wage)=\beta_0+\beta_1 jc+\beta_2 univ+\beta_3 exper+u \tag{4.17}$$

式中，*jc* 为参加两年制大学的年数；*univ* 为参加四年制大学的年数。注意，大专和大学的任意组合都是允许的，包括 $jc=0$ 和 $univ=0$。

我们所关心的假设是，在大专的一年是否比得上在大学的一年。这可表示为

$$\mathrm{H}_0:\ \beta_1=\beta_2 \tag{4.18}$$

137 在 H_0 下，多一年的大专教育与多一年的大学教育，在其他条件不变的情况下，会导致工资同等程度的增加。在多数情况下，我们所关心的对立假设是在大专的一年比不上在大学的一年。这可表示为

$$\mathrm{H}_1:\ \beta_1<\beta_2 \tag{4.19}$$

式（4.18）和式（4.19）中的假设涉及两个参数 β_1 和 β_2，这是一个我们还没有遇到过的情形。我们不能简单地使用 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 单个的 t 统计量来检验 H_0。但从概念上讲，构造一个检验式（4.18）的 t 统计量并没有什么困难。为了构造这个统计量，我们将虚拟假设和对立假设分别重新写成 H_0：$\beta_1-\beta_2=0$ 和 H_1：$\beta_1-\beta_2<0$，这个 t 统计量则基于被估计的差 $\hat{\beta}_1-\hat{\beta}_2$ 是否充分小于零，足以保证拒绝式（4.18）并支持式（4.19）。为了解释估计量中的抽样误差，将这个差值除以其标准误以将其标准化：

$$t=\frac{\hat{\beta}_1-\hat{\beta}_2}{\mathrm{se}(\hat{\beta}_1-\hat{\beta}_2)} \tag{4.20}$$

一旦我们得到式（4.20）中的 t 统计量，检验过程就一如从前。我们选择一个显著性水平，并根据 df 得到一个临界值。因为对立假设具有式（4.19）的形式，所以拒绝法则的形式就是 $t<-c$，其中 c 是从适当的 t 分布中选择的一个正值。或者，我们计算 t 统计量后，再计算 p 值（参见 4.2 节）。

检验两个不同参数之间的相等关系，之所以比检验一个单独的参数 β_j 更困难，仅在于要得到式（4.20）的分母中的标准误。在进行了 OLS 回归之后，分子是很容易得到的。为简明起见，假设使用 $n=285$ 个个人数据得到了如下方程：

$$\begin{aligned}\log(\widehat{wage})&=1.43+0.098jc+0.124univ+0.019exper\\&\quad\ (0.27)\ (0.031)\quad\ (0.035)\qquad\ (0.008)\\&\quad\ n=285,\ R^2=0.243\end{aligned} \tag{4.21}$$

从式（4.21）显然可知，jc 和 $univ$ 对工资都具有经济和统计上的显著影响。这当然有意义，但我们更关心的是，检验这两个系数的估计差值是否统计显著。由于二者之差被估计为 $\hat{\beta}_1-\hat{\beta}_2=-0.026$，读一年大专的回报比读一年大学的回报约低 2.6 个百分点。从经济上讲，这并不是一个无关紧要的差值。-0.026 的差值是式（4.20）中 t 统计量的分子。

不幸的是，方程（4.21）中的回归结果并没有包括得到 $\hat{\beta}_1-\hat{\beta}_2$ 的标准误所需要的足够信息。有人恐怕禁不住要认为 $\mathrm{sd}(\hat{\beta}_1-\hat{\beta}_2)=\mathrm{se}(\hat{\beta}_1)-\mathrm{se}(\hat{\beta}_2)$，但由于在目前这个例子中，$\mathrm{se}(\hat{\beta}_1)-\mathrm{se}(\hat{\beta}_2)=-0.038$，所以毫无意义。由于标准误是标准差的估计值，所以它们肯定总是正值。尽管差值 $\hat{\beta}_1-\hat{\beta}_2$ 的标准误确实取决于 $\mathrm{se}(\hat{\beta}_1)$ 和 $\mathrm{se}(\hat{\beta}_2)$，但其方式多少有些复杂。为了找到 $\mathrm{se}(\hat{\beta}_1-\hat{\beta}_2)$，我们首先要得到这个差值的方差。利用附录 B 中有关方差的结论，我们有

138 $$\mathrm{Var}(\hat{\beta}_1 - \hat{\beta}_2) = \mathrm{Var}(\hat{\beta}_1) + \mathrm{Var}(\hat{\beta}_2) - 2\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) \tag{4.22}$$

仔细观察，这两个方差如何加到一起，然后再减掉 2 倍的协方差。$\hat{\beta}_1 - \hat{\beta}_2$ 的标准差刚好等于式(4.22)的平方根，而且因为$[\mathrm{se}(\hat{\beta}_1)]^2$ 和$[\mathrm{se}(\hat{\beta}_2)]^2$ 分别是$\mathrm{Var}(\hat{\beta}_1)$和 $\mathrm{Var}(\hat{\beta}_2)$的无偏估计量，所以我们有

$$\mathrm{se}(\hat{\beta}_1 - \hat{\beta}_2) = \{[\mathrm{se}(\hat{\beta}_1)]^2 + [\mathrm{se}(\hat{\beta}_2)]^2 - 2s_{12}\}^{1/2} \tag{4.23}$$

式中，s_{12}为$\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2)$ 的一个估计值。我们还没有给出一个求$\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2)$的公式。有些回归软件包的特点是给出 s_{12}，在这种情况下，就可以计算式(4.23) 中的标准误，然后计算式（4.20）中的 t 统计量。附录 E 告诉我们如何使用矩阵代数来求。

我们这里给出的另一种计算方法则简单得多，也不太可能导致什么错误，并能很方便地应用于各种问题。不是试图通过式（4.23）来计算$\mathrm{se}(\hat{\beta}_1 - \hat{\beta}_2)$，而是估计一个能直接给出我们所关心的标准误的不同模型，这要容易得多。将 β_1 和 β_2 之差定义为一个新参数 $\theta_1 = \beta_1 - \beta_2$，于是我们想检验

$$\mathrm{H}_0: \theta_1 = 0 \quad 对 \quad \mathrm{H}_1: \theta_1 < 0 \tag{4.24}$$

式(4.20)中的 t 统计量用$\hat{\theta}_1$ 表示刚好就是$t = \hat{\theta}_1/\mathrm{se}(\hat{\theta}_1)$。关键是如何找到$\mathrm{se}(\hat{\theta}_1)$。

我们的做法是，通过改写模型，以使 θ_1 直接作为方程中一个自变量的系数而出现。由于 $\theta_1 = \beta_1 - \beta_2$，所以可以写成 $\beta_1 = \theta_1 + \beta_2$。代入式(4.17)并重新整理，则给出方程

$$\begin{aligned}\log(wage) &= \beta_0 + (\theta_1 + \beta_2)jc + \beta_2 univ + \beta_3 exper + u \qquad (4.25)\\ &= \beta_0 + \theta_1 jc + \beta_2(jc + univ) + \beta_3 exper + u\end{aligned}$$

主要的发现是，我们在假设检验中所关心的参数 θ_1 现在乘上了变量 jc。截距仍是 β_0，而 $exper$ 的系数仍是 β_3。更重要的是，与 β_2 相乘的是一个新变量，即 $jc + univ$。因此，如果我们想直接估计 θ_1，并得到 $\hat{\theta}_1$ 的标准误，那我们就必须构造一个新变量 $jc + univ$，并在回归模型中用它来取代 $univ$。在本例中，这个新变量有一个自然的解释：高等教育的总年数，所以定义 $totcoll = jc + univ$，并将式（4.25）写成

$$\log(wage) = \beta_0 + \theta_1 jc + \beta_2 totcoll + \beta_3 exper + u \tag{4.26}$$

参数 β_1 已从模型中消失，而 θ_1 却明确地出现在回归方程中。这个模型实际上只是原模型的一个不同写法。我们定义这个新模型的惟一理由是，当我们估计这个新方程时，jc 的系数就是$\hat{\theta}_1$，更重要的是，$\mathrm{se}(\hat{\theta}_1)$会与系数估计值一起报告出来。我们想知道的 t 统计量就是回归软件包对变量 jc（不是 $totcoll$）报告的 t 统计量。

139 当我们用前面用过的 285 个观测进行估计时，结果就是

$$\log(\widehat{wage}) = 1.43 - 0.026jc + 0.124totcoll + 0.019exper$$

$$(0.27)\ (0.018)\ \ (0.035)\qquad\ (0.008)\qquad\qquad (4.27)$$

$$n = 285,\ R^2 = 0.243$$

在这个方程中，不能从方程(4.21)得到的惟一数字就是估计值 -0.026 的标准误0.018。检验式（4.18）的 t 统计量是 $-0.026/0.018 = -1.44$。相对单侧对立假设式（4.19），p 值约为0.075，所以有证据（但不是很强）拒绝式（4.18）。

截距与 *exper* 的斜率估计值及其标准误都与式（4.21）中的结果相同。这一点一定成立，而且它还为检查变形后的方程是否被正确地估计提供了一种方法。新变量 *totcoll* 的系数与式（4.21）中 *univ* 的系数相同，标准误也一样。通过比较（4.17）和式（4.25），我们知道一定是这样。

对 $\theta_1 = \beta_1 - \beta_2$ 计算一个95%的置信区间相当容易。使用标准正态的近似值，可以像通常一样得到这个CI：$\hat{\theta}_1 \pm 1.96se(\hat{\theta}_1)$，这里也就是 -0.026 ± 0.035。

重写模型使之包含我们所关心参数的做法，在各种情况下都能奏效，而且易于实施。（至于其他的例子，可参见习题4.12和习题4.14。）

4.5　对多个线性约束的检验：F 检验

与任意一个OLS系数相联系的 t 统计量，都可以用来检验总体中对应的未知参数是否等于某给定常数（通常但并不总是零）。我们刚刚说明，如何利用变量代换通过重新整理方程而进行回归，得以检验关于 β_j 的一个线性组合。但到目前为止，我们只讨论了涉及单一个约束的假设。我们经常还想检验关于基本参数 β_0，β_1，…，β_k 的多重假设。我们从检验一组自变量是否对因变量都没有影响这个首要问题开始。

对排除性约束的检验

我们已经知道如何检验某一自变量是否对因变量没有偏效应：使用 t 统计量。现在，我们想检验一组自变量是否对因变量没有影响。更准确地说，虚拟假设是，在控制了一些变量之后，余下的那些变量对 y 没有任何影响。

为了阐明检验一组变量的显著性是有用的，让我们考虑如下解释主要俱乐部中棒球运动员薪水的模型：

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + u \qquad (4.28)$$

140　式中，*salary* 为1993年的总薪水；*years* 为加入俱乐部的年资；*gamesyr* 为平均每年的比赛次数；*bavg* 为平均职业击球次数（比方说 $bavg = 250$）；

hrunsyr 为平均每年的本垒打次数；*rbisyr* 为每年的击球跑垒得分。假设我们想检验的虚拟假设是，一旦控制了加入俱乐部的年资和每年的比赛次数，度量球员表现的统计指标（*bavg*，*hrunsyr* 和 *rbisyr*）对薪水没有影响。从本质上讲，虚拟假设所表达的是由棒球统计指标度量的生产力对薪水没有影响。

虚拟假设可以使用模型的参数表示为

$$\mathrm{H}_0: \beta_3=0, \ \beta_4=0, \ \beta_5=0 \tag{4.29}$$

虚拟假设（4.29）由三个**排除性约束**（exclusion restrictions）构成：如果假设（4.29）是正确的，那么在控制了 *years* 和 *gamesyr* 之后，*bavg*，*hrunsyr* 和 *rbisyr* 对 log（*salary*）没有影响，就应该把它们从模型中排除。这是**多重约束**（multiple restrictions）的一个例子，因为我们对式（4.28）中的参数施加了不止一个约束；以后我们还将看到多重约束的更一般的例子。对多重约束进行的检验被称为**多重假设检验**（multiple hypotheses test）或**联合假设检验**（joint hypotheses test）。

式（4.29）的对立假设应该是什么呢？如果我们认为是“即使控制了加入俱乐部年数和每年的比赛数，表现的统计指标也有影响”，那么合适的对立假设也就是

$$\mathrm{H}_1: \mathrm{H}_0 \text{ 不正确} \tag{4.30}$$

如果 β_3，β_4，或 β_5 中至少有一个异于零，那么对立假设（4.30）就成立。（可以是任何一个或全部异于零。）我们这里构造的检验是为了侦查是否有违背 H_0 的迹象。尽管也可以当对立假设采用诸如 H_1：$\beta_3>0$，或 $\beta_4>0$，或 $\beta_5>0$之类的形式，但在这种对立假设之下，上述检验不会是最好的检验。我们没有足够的篇幅和统计方面的知识来讨论那种在多个单侧对立假设下更具功效的检验。

我们应该如何针对假设（4.30）检验假设（4.29）呢？人们不禁要想到使用变量 *bavg*，*hrunsyr* 和 *rbisyr* 的 t 统计量，以决定是否每个变量都是个别显著的，并以此来检验假设（4.29）。但这种做法并不合适。一个特定的 t 统计量只能检验一个对其他参数没有任何限制的假设。此外，我们还要费力对付三个结果——每个 t 统计量给出一个结果。怎样才能在 5％的显著性水平上拒绝假设（4.29）呢？应该要求所有这三个或只是其中一个 t 统计量在 5％的显著性水平上显著吗？这些都是难以回答的问题，而幸运的是，我们不必回答这些。而且，分别使用的 t 统计量来检验一个像式（4.29）那样的多重假设，可能极具误导性。我们需要一个能联合检验这些排除性约束的方法。

为了说明这些问题，我们使用 MLB1.RAW 中的数据来估计方程（4.28）。结果是

$$\begin{aligned}\log(salary) = {} & \underset{(0.29)}{11.10} + \underset{(0.012\ 1)}{0.068\ 9}years + \underset{(0.002\ 6)}{0.012\ 6}gamesyr \\ & + \underset{(0.001\ 10)}{0.000\ 98}bavg + \underset{(0.016\ 1)}{0.014\ 4}hrunsyr + \underset{(0.007\ 2)}{0.010\ 8}rbisyr \qquad (4.31) \\ & n = 353,\ \mathrm{SSR} = 183.186,\ R^2 = 0.627\ 8\end{aligned}$$

141 式中，SSR 为残差平方和。(我们以后将用到它。)我们在 SSR 和 R^2 的小数点后都保留了几位，以便以后进行比较。方程(4.31)显示，在 *years* 和 *gamesyr* 都统计显著的同时，针对双侧对立假设，*bavg*，*hrunsyr* 和 *rbisyr* 中没有一个变量在 5% 的显著性水平上具有一个统计显著的 t 统计量。(*rbisyr* 的 t 统计量接近于显著；其双侧对立假设的 p 值是 0.134。)于是，从三个 t 统计量来看，我们不能拒绝 H_0。

但这个结论却是错误的。为了看出这一点，我们必须导出一个对多重约束的检验，而且它的分布是已知且被列成表格的。我们将看到残差平方和为检验多重假设提供了一个很方便的途径。我们还将证明，R^2 如何用于检验排除性约束的特殊情形。

知道式(4.31)中的残差平方和，并不能告诉我们式(4.29)中假设的真伪。但它能告诉我们，当将 *bavg*，*hrunsyr* 和 *rbisyr* 从模型中去掉时 SSR 会增大多少。记住，因为所选择的 OLS 估计值都是为了最小化残差平方和，所以当从模型中去掉变量时，SSR 总是增加的；这是数学上的一个事实。问题是，这种增加**相对**含有这些变量的模型的 SSR 来说，是否大到能足以拒绝虚拟假设。

不含上述三个变量的模型便是

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + u \qquad (4.32)$$

在假设检验的背景下，方程（4.32）是检验（4.29）的**受约束模型**（restricted model）；模型（4.29）则被称为**不受约束模型**（unrestricted model）。受约束模型的参数总比不受约束模型的参数要少一些。

当用 MLB1.RAW 中的数据来估计受约束模型时，我们得到

$$\begin{aligned}\log(salary) = {} & \underset{(0.11)}{11.22} + \underset{(0.012\ 5)}{0.071\ 3}years + \underset{(0.001\ 3)}{0.020\ 2}gamesyr \qquad (4.33) \\ & n = 353,\ \mathrm{SSR} = 198.311,\ R^2 = 0.597\ 1\end{aligned}$$

如我们所想，来自式（4.33）的 SSR 比来自式（4.31）的 SSR 要大些，而受约束模型的 R^2 比不受约束模型的 R^2 要小些。我们需要决定的是，SSR 从不受约束模型到受约束模型的增加（从 183.186～198.311），是否足以拒绝式（4.29）。同所有的检验一样，答案取决于检验的显著性水平。但我们在得到一个在 H_0 下其分布已知且可制成表格的统计量之前，还不可能在一个选定的显著性水平下进行检验。因此，我们需要一种方法，能合并这两个 SSR 中的信息，得到一个在 H_0 下分布已知的检验统计量。

我们不妨针对一般情形推导这个检验统计量，这样做也不会增加多少困难。将具有 k 个自变量的不受约束模型写成

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{4.34}$$

142 不受约束模型中的参数有 $k+1$ 个。（记住要增加一个截距参数。）假设我们有 q 个排除性约束要检验，即虚拟假设表示式（4.34）中有 q 个变量的系数为零。为了记号上的方便，假定这 q 个变量是自变量中的最后 q 个：x_{k-q+1}，…、x_k。（当然变量的顺序是任意的，同时也是不重要的。）虚拟假设便被表示成

$$H_0: \beta_{k-q+1} = 0, \cdots, \beta_k = 0 \tag{4.35}$$

它对模型（4.34）施加了 q 个排除性约束。式（4.35）的对立假设，简单地说，就是式（4.35）是错误的；这就意味着式（4.35）中列出的参数至少有一个异于零。当我们施加在 H_0 下的约束时，就给我们留下了受约束模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-q} x_{k-q} + u \tag{4.36}$$

在这一小节，我们假定受约束模型和不受约束模型都包括截距项，因为在实践中，含有截距项是最经常遇到的情况。

现在轮到检验统计量本身了。前面我们提到，当不受约束模型变到受约束模型时，SSR 的相对增加对检验假设（4.35）而言应该是有意义的。定义 **F 统计量**（F statistic 或 F 比率）为

$$F \equiv \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)} \tag{4.37}$$

式中，SSR_r 为受约束模型的残差平方和；SSR_{ur} 为不受约束模型的残差平方和。

问题 4.4

考虑学生在标准化考试中的个人表现 *score* 与一系列其他变量之间的关系。学校因素包括人均教室面积、学生平均支出、教师的平均薪水和全校总注册人数。其他学生特征变量为家庭收入、母亲的受教育程度、父亲的受教育程度和兄弟姐妹的个数。这个模型是

$$\begin{aligned} score = {} & \beta_0 + \beta_1 classize + \beta_2 expend + \beta_3 tchcomp + \beta_4 enroll \\ & + \beta_5 faminc + \beta_6 motheduc + \beta_7 fatheduc + \beta_8 siblings + u \end{aligned}$$

表述一旦与学校相关的因素得到控制以后学生特征变量对标准化考试的表现没有影响这个虚拟假设。这个例子的 k 和 q 是什么？写出这个模型的受约束型。

你可能立即就注意到了，因为 SSR_r 不可能比 SSR_{ur} 小，所以 F 统计量总是非负（而且几乎总是严格为正）。因此，如果你计算出了一个负的 F 统计量，那就肯定出了什么问题；通常是 F 统计量的分子中两个 SSR 的次序被颠倒了。同时，F 统计量分母中的 SSR 来自**不受约束模型**。最容易记住 SSR 出现方式的办法是，认为 F 度量的是 SSR 从不受约束模型到受约束模

型的相对提高。

F 统计量分子中的 SSR 的差值还要除以 q，q 就是从不受约束模型到受约束模型所施加的约束数（去掉了 q 个自变量）。于是可以写出

$$q = \text{分子自由度} = df_r - df_{ur} \tag{4.38}$$

143 它还表明，q 是受约束模型与不受约束模型的自由度之差。（回忆 df = 观测次数 − 被估计参数的个数。）由于受约束模型参数较少，而每个模型都使用同样的 n 次观测，所以 df_r 总是大于 df_{ur}。

F 统计量分母中的 SSR 要除以不受约束模型的自由度：

$$n - k - 1 = \text{分母自由度} = df_{ur} \tag{4.39}$$

实际上，F 的分母恰好就是不受约束模型中 $\sigma^2 = \mathrm{Var}(u)$ 的一个无偏估计量。

在具体应用中，计算 F 统计量比之于艰难完成的一般情形的描述要容易些，因为在描述一般情形时使用了许多繁琐的符号。我们首先找出不受约束模型的自由度 df_{ur}。然后我们数一下在受约束模型中排除了多少个变量，即 q。每个 OLS 回归都会报告 SSR，所以形成 F 统计量就很简单。

在大棒球俱乐部薪水的回归模型中，$n = 353$，完整模型（4.28）包含了 6 个变量。于是，$n - k - 1 = df_{ur} = 353 - 6 = 347$。受约束模型（4.32）比式（4.28）少包括 3 个变量，所以 $q = 3$。因此，我们具有计算 F 统计量所需要的所有要素；在我们知道 F 统计量用来干什么之前，我们暂不计算它。

为了使用 F 统计量，我们必须知道它在虚拟假设下的抽样分布，以选定临界值和拒绝法则。可以证明，在 H_0 下（并假设 CLM 假定成立），F 统计量服从自由度为 $(q, n-k-1)$ 的 F 随机变量的分布。我们把它写成

$$F \sim F_{q,\, n-k-1}$$

$F_{q,\, n-k-1}$ 分布，已制成表格，可在统计表中查阅（见表 G.3），更重要的是，在统计软件中也可获得。

由于推导 F 分布所需要的数学十分复杂，所以我们不去推导它的分布。大致说来，可以证明，式（4.37）实际上是两个独立的 χ 平方随机变量分别除以其自由度后的比率。分子 χ 平方随机变量的自由度为 q，分母 χ 平方随机变量的自由度则为 $n - k - 1$。这就是一个 F 分布随机变量的定义（参见附录 B）。

从 F 的定义能相当明显地看出，当 F 充分“大”时，我们将拒绝 H_0 而支持 H_1。需要多大则取决于我们选择的显著性水平。假设我们决定在 5% 的水平上进行检验。令 c 为 $F_{q,\, n-k-1}$ 分布的第 95 个百分位，这个临界值取决于 q（分子自由度）和 $n - k - 1$（分母自由度）。重要的是弄清楚分子和分母的自由度。

F 分布 10%，5% 和 1% 的临界值在表 G.3 中给出。拒绝法则很简单。一旦得到了 c，如果

$$F > c \tag{4.40}$$

144 我们就在所选定的显著性水平上拒绝 H_0 而支持 H_1。对于5%的显著性水平，$q=3$ 和 $n-k-1=60$，那么临界值就是 $c=2.76$。如果计算出来的 F 统计量的值超过2.76，我们就在5%的显著性水平上拒绝 H_0。图4.7画出了5%的显著性水平和拒绝域。对于同样的自由度，1%水平的临界值则是4.13。

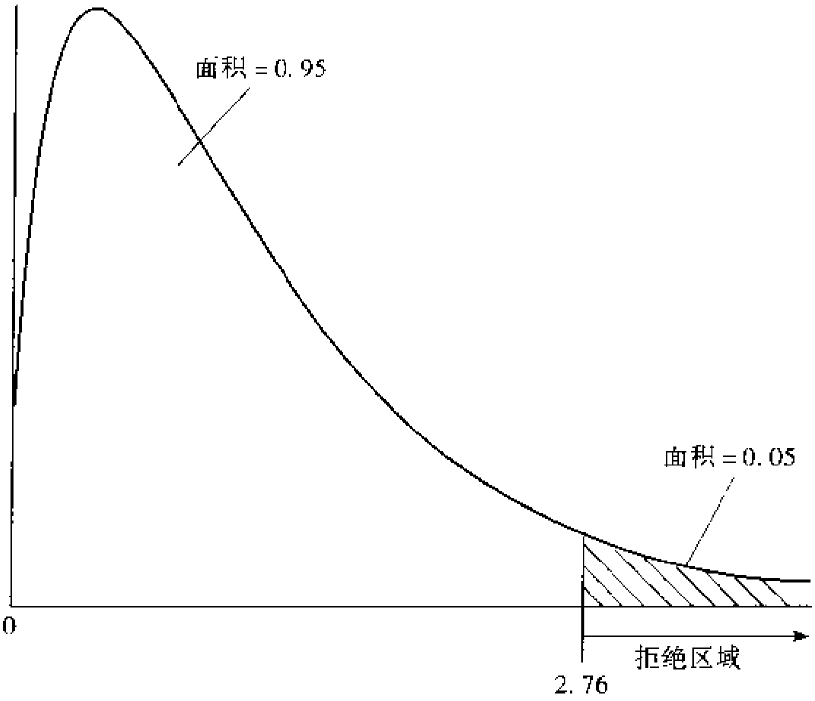


图 4.7　在一个 $F_{3,60}$ 分布中 5%的临界值和拒绝域

在大多数应用中，分子自由度（q）都比分母自由度($n-k-1$)要小得多。在 $n-k-1$ 不大的情况中，可能因为不能准确估计虚拟模型中的参数，所以应用可能会不太成功。当分母自由度达到约120时，F 分布对它就不再敏感了。（这完全就像 t 分布随着 df 的变大而能很好地由标准正态分布来近似一样。）因此，分布表中专门有一个 $df=\infty$ 的条目，这就是我们在大样本（于是 $n-k-1$ 也大）情况下所用的临界值。对于分子自由度极大的情况，类似的论断也成立，但在实际应用中这种情况却很少发生。

如果拒绝 H_0，那么我们就说 $x_{k-q+1}, \cdots, x_k$ 在适当的显著性水平上是
145 **联合统计显著**（jointly statistically significant）的（或简单地说**联合显著的**）。单靠这一个检验还不足以告诉我们哪个变量对 y 具有影响；它们或许都影响 y，或许只有一个影响 y。如果虚拟假设未被拒绝，那么这些变量就是**联合不显著**（jointly insignificant）的，这就为我们将它们从模型中去掉提供了根据。

对于大棒球俱乐部的例子，分子自由度为3，分母自由度为347，显著性水平为5%的临界值是2.60，而1%的临界值是3.78；如果 F 大于2.60，就在5%的水平上拒绝虚拟假设。

现在回过头来检验我们在本节之初的假设：在控制了 *years* 和 *gamesyr* 后，变量 *bavg*、*hrunsyr* 和 *rbisyr* 对运动员的薪水没有影响。实际上，最简

单的办法是，先计算$(\mathrm{SSR}_r-\mathrm{SSR}_{ur})/\mathrm{SSR}_{ur}$，再将结果乘以$(n-k-1)/q$；把公式表述成式（4.37）那样的原因，是使我们更容易弄清楚分子和分母的自由度。使用式（4.31）和式（4.33）中的SSR，我们得到

$$F=\frac{(198.311-183.186)}{183.186}\cdot\frac{347}{3}\approx 9.55$$

这个数字远大于自由度为3和347的F分布在显著性水平为1%的临界值，所以，为稳妥起见，我们拒绝*bavg*，*hrunsyr*和*rbisyr*对薪水没有影响的假设。

鉴于三个变量的t统计量都不显著，联合检验的结果看上去有些令人吃惊。事实情况是，*hrunsyr*和*rbisyr*两个变量高度相关，而这种多重共线性则使得我们难以发现每个变量的偏效应；后者反映在单个t统计量上。F统计量检验这三个变量（包括*bavg*）是否联合显著，而*hrunsyr*和*rbisyr*之间的多重共线性对检验这个联合假设而言就远没有那么重要了。在习题4.16中，要求你重新估计去掉*rbisyr*之后的模型，在这种情况下，*hrunsyr*就十分显著。将*hrunsyr*从模型中去掉后，情况也是一样的。

F统计量通常对检验一组变量的排除有用处，特别是在其中的变量高度相关的时候。例如，我们想检验企业的业绩是否影响总经理的薪水。有许多度量企业业绩的方法，可能事先并不清楚哪种度量方法是最重要的。既然对业绩的度量可能高度相关，那么，由于多重共线性，希望发现个别显著的度量就可能要求过高。但F检验则可用来决定一组企业业绩变量是否会影响薪水。

F统计量和t统计量之间的关系

在本节我们已经看到，如何用F统计量来检验模型中是否应该包括某一组变量。如果我们用F统计量去检验单一个自变量的显著性，结果会怎么样呢？前面的论述当然没有排除这种可能情形。比如，我们可以取虚拟假设为H_0：$\beta_k=0$和$q=1$（以便检验x_k可从模型中去掉的单一变量排除约束）。在4.2节我们知道，β_k的t统计量可用来检验这个假设。那么问题是，
146 对于单一个系数是否有两种截然不同的假设检验方法。回答是否定的。可以证明，检验单一个变量之排除性的F统计量，等于对应t统计量的平方。因为t_{n-k-1}^2具有$F_{1,n-k-1}$分布，所以在双侧对立假设下，这两种方法得到完全一样的结果。由于t统计量可用来检验单侧对立假设，所以它对于检验单一个参数假设就更灵活。还因为t统计量比F统计量更容易获得，所以实在没有理由去使用F统计量对单个参数假设进行检验。

F 统计量的 R-平方型

多数应用表明，使用受约束模型和不受约束模型的 R 平方来计算 F 统计量将更方便。原因之一在于，R 平方必定介于 0 和 1 之间，而 SSR 则在很大程度上依赖于度量单位，使得基于 SSR 的计算繁冗。利用 $\mathrm{SSR}_r = \mathrm{SST}(1-R_r^2)$ 和 $\mathrm{SSR}_{ur} = \mathrm{SST}(1-R_{ur}^2)$，代入式（4.37），得到

$$F=\frac{(R_{ur}^2-R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)} \tag{4.41}$$

（注意消掉了 SSR 项）。此即所谓的 **R-平方型 F 统计量**（R-squared form of the F statistic）。

由于 R-平方在几乎所有的回归中都会报告（但 SSR 却没有），所以，使用受约束模型和不受约束模型的 R-平方来检验一些变量的排除就很容易。应该特别注意分子中 R-平方的顺序：先出现无约束模型的 R-平方［与式（4.37）中 SSR 的顺序相反］。因为 $R_{ur}^2 > R_r^2$，所以再次表明 F 总是正的。

在使用 R-平方型统计量检验一组变量的排除时，重要的是，在将 R-平方放到式（4.41）中之前，不能再将它平方；平方运算已经做过了。所有的回归都报告 R^2，把这些数字都直接代入式（4.41）就是。如棒球运动员薪水一例，我们可以使用式（4.41）得到 F 统计量：

$$F=\frac{(0.6278-0.5971)}{1-0.6278}\cdot\frac{347}{3}\approx 9.54$$

与我们前面得到的结论十分接近。（差别在于四舍五入的误差。）

例 4.9　孩子出生体重方程中父母的受教育水平

作为计算 F 统计量的另一个例子，考虑如下用各种因素解释婴儿出生体重的模型：

$$\begin{aligned} bwght &= \beta_0+\beta_1 cigs+\beta_2 parity+\beta_3 faminc \\ &\quad +\beta_4 matheduc+\beta_5 fatheduc+u \end{aligned} \tag{4.42}$$

147 式中，*bwght* 为以磅为单位的出生体重；*cigs* 为母亲怀孕期间平均每天吸烟的数量；*parity* 为这个孩子在子女中的排行；*faminc* 为家庭年收入；*motheduc* 为母亲受教育年数；*fatheduc* 为父亲受教育年数。我们要检验的虚拟假设是，在控制了 *cigs*，*parity* 和 *faminc* 后，父母的受教育水平对孩子出生体重没有影响。表示为 H_0：$\beta_4=0$，$\beta_5=0$，所以有 $q=2$ 个排除性约束要检验。在不受约束模型（4.42）中有 $k+1=6$ 个参数，所以在不受约束模型中的 df 是 $n-6$，其中，n 为样本容量。

我们将使用 BWGHT.RAW 中的数据来检验这个假设。尽管这个数据集包含了 1 388 个婴儿的信息，但在清点检验虚拟假设所用到的观测数据时仍

须十分小心。结果，样本中有197个婴儿至少缺失变量 *motheduc* 和 *fatheduc* 中的一个信息；在估计不受约束模型时，这些观测都不能被包含进来。因而我们实际上只有1 191个观测值，所以在不受约束模型中也就只有1 191 − 6 = 1 185个 *df*。在估计受约束模型时，我们一定要使用这同样的1 191个观测（而不是可供使用的全部1 388个观测）。一般而言，在估计受约束模型以计算 *F* 检验时，我们必须使用与估计不受约束模型同样的观测，否则这个检验就无效。在没有缺失数据的情况下，这当然不成什么问题

分子的 *df* 为2，分母的 *df* 为1 185；查表G.3，知5%的临界值是3.0。为了节省篇幅，我们就不报告计算结果，而只是给出 *R*-平方的值。对于完整的模型，*R*-平方是 $R_{ur}^2 = 0.0387$；将 *motheduc* 和 *fatheduc* 从回归中去掉后，*R*-平方下降到 $R_r^2 = 0.0364$。于是，*F* 统计量是 $F = [(0.0387 - 0.0364)/(1 - 0.0387)](1\,185/2) = 1.42$；因为它远远低于5%的临界值，所以我们不能拒绝 H_0。换句话说，*motheduc* 和 *fatheduc* 在婴儿体重方程中是联合不显著的。

计算 *F* 检验的 *p* 值

p 值对报告 *F* 检验的结果特别有用。由于 *F* 分布取决于分子和分母的 *df*，所以只是看一下 *F* 统计量的值或一二个临界值，对拒绝虚拟假设之证据的强弱很难有直观感觉。

在 *F* 检验的背景下，*p* 值被定义为

138

$$p\text{值} = P(\mathscr{F} > F) \tag{4.43}$$

式中，为了强调，我们令 $\mathscr{F}$ 表示一个自由度为（q, $n-k-1$）的 *F* 随机变量，*F* 是检验统计量的实际值。*p* 值与 *t* 统计量的 *p* 值具有相同的解释：观察到 *F* 值至少和我们在给定虚拟假设是正确时所观察到的 *F* 值一样大的概率。很小的 *p* 值就是拒绝 H_0 的证据。比如，*p* 值 = 0.016 意味着观察到 *F* 值至少和虚拟假设正确时所观察到的 *F* 值一样大的机会是1.6%；我们在这种情况下通常会拒绝 H_0。如果 *p* 值 = 0.314，那么，观察到 *F* 统计量的值至少和我们在虚拟假设正确时所观察到的 *F* 值一样大的机会是31.4%。大多数人认为，要拒绝 H_0 的话，这还是相当弱的证据。

如同 *t* 检验一样，一旦计算了 *p* 值，*F* 检验可以在任何显著性水平下进行。比方说，如果 *p* 值 = 0.024，我们将在5%的显著性水平上拒绝 H_0，但在1%的显著性水平上则不能拒绝。

问题4.5 用ATTEND.RAW中的数据去估计两个方程

$$\widehat{atndrte} = 47.13 + 13.37\, priGPA$$
$$(2.87)\quad(1.09)$$

$$n = 680,\ R^2 = 0.183$$

和

$$\widehat{atndrte} = 75.70 + 17.26 priGPA - 1.72 ACT$$
$$(3.88)\ (1.08)\qquad\qquad (?)$$
$$n = 680,\ R^2 = 0.291$$

标准误与通常一样放在括号中；第二个方程缺少 ACT 的标准误。ACT 系数的 t 统计量是什么？[提示：首先计算 ACT 显著性的 F 统计量。]

例 4.9 中 F 检验的 p 值是 0.238，所以 $\beta_{motheduc}$ 和 $\beta_{fatheduc}$ 都是零的虚拟假设即使在 20% 的显著性水平上也不能拒绝。

许多计量经济软件包都具有检验多重排除约束的内在性能。与用手计算这些统计量相比，这些软件包有几个优点：我们出错的可能性减少，p 值被自动计算出来，像例 4.9 那样的数据缺失问题，也无须我们花任何功夫就得到解决。

回归整体显著性的 F 统计量

大多数回归软件包还对一组特定的排除约束进行了例行检验。无论是哪个模型，这些约束都具有相同的解释。在含有 k 个自变量的模型中，我们可以把虚拟假设写成

$$H_0:\ x_1,\ x_2,\ \cdots,\ x_k \text{ 无助于解释 } y$$

这个虚拟假设从某个方面来讲太过悲观。它认为解释变量中没有一个能影响 y。用参数表示，这个虚拟假设就是所有的斜率参数都是零：

$$H_0:\ \beta_1 = \beta_2 = \cdots = \beta_k = 0 \tag{4.44}$$

而对立假设是至少有一个 β_j 异于零。另一种表述虚拟假设的有用方法是，H_0：$E(y|x_1,x_2,\cdots,x_k) = E(y)$，所以，知道 x_1，x_2，…，x_k 的值不会影响 y 的期望值。

在式（4.44）中有 k 个约束，当我们施加这些约束时，得到受约束模型

$$y = \beta_0 + u \tag{4.45}$$

所有的自变量都已经从方程中去掉了。现在，估计式（4.45）的 R-平方为零；因为没有解释变量，所以 y 中的变异一点都没有得到解释。因此，检验式（4.44）的 F 统计量可写成

$$\frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{4.46}$$

式中，R^2 就是 y 对 x_1，x_2，…，x_k 回归的通常 R-平方。

149 大多数回归软件包都自动报告式（4.46）中的 F 统计量，使得利用这个统计量去检验一般的排除性约束很有诱惑力。你一定要抵制住这种诱惑。

通常使用式（4.41）中的 F 统计量来检验一般的排除性约束，它取决于受约束模型和不受约束模型的 R-平方。式（4.46）的特殊形式，只有在检验所有自变量的联合排除时才有效。有时称之为检验回归的**整体显著性**（overall significance of the regression）。

如果我们不能拒绝式（4.44），就没有证据表明有某个自变量能有助于解释 y。这通常意味着我们必须寻找其他的变量来解释 y。对于例 4.9，检验式（4.44）的 F 统计量在 $k=5$ 和 $n-k-1=1\ 185$ 时约为 9.55。p 值在小数点后面四位都是零，所以强有力地拒绝了式（4.44）。因此，我们断定，*bwght* 方程中的变量**确实**能解释 *bwght* 中的某些变异。被解释的变异并不多，只有 3.87%。但看上去很小的 R-平方，却导致高度显著的 F 统计量。这就解释了我们为什么要计算 F 统计量来检验联合显著性，而不是仅仅看一下 R-平方的大小。

有时候，所有自变量联合不显著之假设的 F 统计量也是研究的焦点。习题 4.10 就要求你利用股票回报数据，去检验股票四年后的回报能否仅由期初所拥有的信息来预测。在**有效市场假设**（efficient markets hypothesis）下，回报应该是不可预测的；虚拟假设刚好就是式（4.44）。

检验一般的线性约束

检验排除性约束到目前为止仍是 F 统计量最重要的应用。但有时候，一种理论所蕴涵的约束比仅仅排除某些自变量更为复杂。这仍可以直接使用 F 统计量进行检验。

举一个例子，考虑如下方程：

$$
\begin{aligned}
\log(price) = \beta_0 + \beta_1\log(assess) + \beta_2\log(lotsize) \\
+ \beta_3\log(sqrft) + \beta_4 bdrms + u
\end{aligned}
\qquad (4.47)
$$

式中，*price* 为住房价格；*assess* 为评估的住房价值（在房子售出以前）；*lotsize* 为以英尺为单位的整体尺寸；*sqrft* 为平方英尺数；*bdrms* 为卧室数。现在，假设我们想检验评估的住房价值是不是一个理性的定价。如果是这样，那么 *assess* 变化 1%，将伴随着 *price* 变化 1%，即 $\beta_1=1$。此外，一旦控制了评估价值 *assess* 后，*lotsize*，*sqrft* 和 *bdrms* 应该无助于解释 log（*price*）。总而言之，这些假设可表述成

$$
\mathrm{H}_0: \beta_1 = , \beta_2 = 0, \beta_3 = 0, \beta_4 = 0 \qquad (4.48)
$$

这里有四个约束要检验，有三个是排除性约束，但 $\beta_1=1$ 不是。我们如何使用 F 统计量来检验这个假设呢？

像在排除性约束情形中一样，我们先估计不受约束模型，这里就是式 150 （4.47），然后施加式（4.48）中的约束，得到受约束模型。第二步多少有些微妙。但我们所做的一切就是代入约束。如果我们将式（4.47）写成

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+u \tag{4.49}$$

那么受约束模型就是 $y=\beta_1+x_1+u$。现在，为了施加 x_1 的系数为 1 的约束，我们必须估计如下模型：

$$y-x_1=\beta_0+u \tag{4.50}$$

这只是一个具有一个截距（β_0）的模型，但与式（4.49）中的因变量不同。计算 F 统计量的程序是一样的：估计式（4.50），得到 SSR（SSR_r），并将它与来自式（4.49）的不受约束模型一起用于 F 统计量（4.37）。我们在检验 4 个约束，所以在不受约束模型中有 $n-5$ 个 df。F 统计量也就是 $[(\text{SSR}_r-\text{SSR}_{ur})/\text{SSR}_{ur}][(n-5)/4]$。

在使用一个数据集来说明这个检验之前，我们必须强调一点：由于式(4.50)中的因变量不同于式(4.49)中的因变量，所以在这个例子中不能使用 F 统计量的 R-平方型。这意味着两个回归的总平方和将不同，而式(4.41)也就不再等价于式(4.37)。作为一个一般规律，如果在进行受约束回归时需要一个不同的因变量，那就应该使用 F 统计量的 SSR 型。

使用 HPRICE1.RAW 中的数据估计的不受约束模型是

$$\begin{aligned}
\log(\widehat{price}) = {}& -0.034 + 1.043\log(assess) + 0.007\,4\log(lotsize)\\
& (0.972)\quad (0.151)\qquad\qquad (0.038\,6)\\
& -0.103\,2\log(sqrft) + 0.033\,8bdrms\\
& (0.138\,4)\qquad\qquad\qquad (0.022\,1)\\
& n=88,\ \text{SSR}=1.822,\ R^2=0.773
\end{aligned}$$

如果分别用 t 统计量检验式（4.48）中的每一个假设，那我们一个也不能拒绝。但“评估是理性的”是一个联合假设，所以应该联合检验这些约束。结果受约束模型的 SSR 是 $\text{SSR}_r=1.880$，所以 F 统计量便是$[(1.880-1.822)/1.822](83/4)=0.661$。自由度为（4，83）的 F 分布，在显著性水平为 5% 时的临界值约为 2.50，所以我们不能拒绝 H_0。实际上，没有证据拒绝评估值是理性的这个假设。

4.6 报告回归结果

在本章结束之际，我们就如何对相对复杂的经验研究报告多元回归的结果给出一些指导性原则。这会让你知道如何阅读应用社会科学中发表的作品，同时也为你写自己的经验论文做准备。在本书其余部分，我们还将通过报告各种例子的结果来补充这个专题，但其中的许多要点现在就可以给出。

当然，所估计的 OLS 系数估计值总应该报告。对于分析中的关键变量，你应该对所估计的系数作出解释（这通常要求知道这些变量的度量单位）。

151 比如，这个估计值是不是一个弹性，还是有需要解释的其他理解？应该对关键变量估计值的经济或实际重要性加以讨论。

标准误总是应该与所估计的系数一起包括进来。有些作者更喜欢报告 t 统计量而不是标准误（通常只是 t 统计量的绝对值）。尽管这并没有什么过错，但报告标准误还是要好一些。首先，标准误迫使我们认真考虑被检验的虚拟假设；虚拟假设并非总是总体参数为零。其次，有了标准误，计算置信区间就更容易。

回归的 R-平方也总应该包括进来。我们已经看到，除了提供拟合优度的一种度量外，它还使计算排除性约束的 F 统计量相对简单。报告残差平方和及回归标准误有时也是一个好主意，但并非至关重要。估计任何一个方程的观测次数也应该出现在估计出来的方程中。

如果只估计少数几个模型，像我们到目前为止所做的那样，结果就可以总结在方程形式中。但在许多论文中，几个方程由许多不同的自变量集来估计。我们可能要对不同的人群估计同一个方程，或者我们的方程所解释的是不同的因变量。对于这些情形，最好将结果归纳在一个或多个表格中。表中应该清楚标明因变量，而自变量则应该列在第一列。标准误（或 t 统计量）可放在估计值下面的括号中。

例 4.10　教师的薪水—养老金之间的替换关系

令 *totcomp* 表示一个教师全年平均的总报酬，包括薪水和各种附加福利（养老金、健康保险等）。把标准工资方程加以扩展，总报酬应该是生产力及其他特征的一个函数。如标准做法那样，我们使用对数形式：

$$\log(totcomp)=f\,(\text{生产力特征,其他特征})$$

式中，$f(\cdot)$是某种（尚未设定的）函数。记

$$totcomp = salary + benefits = salary(1 + benefits/salary)$$

这个方程表明，总报酬是 *salary* 和 $1+b/s$ 两项之积，b/s 是“福利—薪水比”的简写。将这个方程取对数就得到 $\log(totcomp)=\log(salary)+\log(1+b/s)$。现在，对于“小”的 b/s，$\log(1+b/s)\approx b/s$；我们将使用这个近似。这就形成了一个计量模型

$$\log(salary)=\beta_0+\beta_1(b/s)+\text{其他因素}$$

于是，检验工资—福利之间的替换关系，就等同于检验 H_0：$\beta_1=-1$，对立假设是：H_1：$\beta_1\neq-1$。

我们使用 MEAP93.RAW 中的数据去检验这个假设。这些数据是在学校
152 层次上的平均值，而且我们没有观测到很多能影响总报酬的其他因素。我们要控制的因素将包括学校规模（*enroll*），平均每千名学生的教工数（*staff*），以及对诸如学校辍学率和毕业率等的度量。样本中平均的 b/s 约为 0.205，而其最大值为 0.450。

所估计的方程在表 4.1 中给出，其中标准误放在系数估计值下面的括号

中。关键变量是福利—薪水比 b/s。

问题 4.6 加入 *droprate* 和 *gadrate* 后，如何影响薪水—福利替换关系的估计值？这些变量在 5% 的显著性水平上是联合统计显著的吗？在 10% 的水平上呢？

从表 4.1 的第一列我们看到，不控制任何其他因素，b/s 的 OLS 系数为 -0.825。检验虚拟假设 $H_0: \beta_1 = -1$ 的 t 统计量是 $t = (-0.825+1)/0.200 \approx 0.875$，所以简单回归不能拒绝 H_0。在引入学校规模和教师规模（大致刻画了每个教师所教学生数）后，b/s 的系数估计值就变为 -0.605。现在检验 $\beta_1 = -1$的 t 统计量约为 2.39；因此，H_0 相对双侧对立假设而在 5% 的显著性水平上被拒绝。变量 $\log(enroll)$ 和 $\log(staff)$ 都是极为统计显著的。

表 4.1　　检验薪水—福利之间的替换关系

自变量	因变量：$\log(salary)$ (1)	 (2)	 (3)
b/s	−0.825	0.605	0.589
	(0.200)	(0.165)	(0.165)
$\log(enroll)$	—	−0.087 4	−0.088 1
		(0.007 3)	(0.007 3)
$\log(staff)$	—	−0.222	−0.218
		(0.050)	(0.050)
droprate	—	—	−0.000 28
			(0.001 61)
gadrate	—	—	−0.000 97
			(0.000 66)
截　　距	10.523	10.884	10.738
	(0.042)	(0.252)	(0.258)
观测次数	408	408	408
R-平方	0.040	0.353	0.361

▶ 小　结

153 我们在本章已探讨了统计推断这个十分重要的专题，使我们能从一个随机样本来对总体模型作出一些推断。现将要点总结如下：

1. 在经典线性模型假定 MLR.1～MLR.6 下，OLS 估计量是正态分布的。

2. 在 CLM 假定下，t 统计量在虚拟假设条件下服从 t 分布。

3. 我们利用 t 统计量针对单侧或双侧对立假设对单一个参数进行假设检验（分别使用单尾或双尾检验）。最常见的虚拟假设是 H_0：$\beta_j=0$，但我们有时也想在 H_0 下检验 β_j 的其他值。

4. 在经典线性假设检验中，首先选择一个显著性水平，它与 df 和对立假设一起决定了临界值，然后就将 t 统计量与这个临界值相比较。对 t 检验计算 p 值（拒绝虚拟假设的最小显著性水平）更有意义，从而我们能在任何一个显著性水平下对虚拟假设进行检验。

5. 在 CLM 假定下，可以对每个 β_j 构造一个置信区间。这些 CI 可以用来对任何一个关于 β_j 的虚拟假设进行相对于一个双侧对立假设的检验。

6. 通过改写模型以（直接）包含我们所关注的参数，总可以检验那些涉及不止一个 β_j 的单个假设检验。随之，就可以使用标准的 t 统计量。

7. F 统计量可用于检验多重排除约束，而且有两种不同的检验形式。一种是基于受约束模型和不受约束模型中的 SSR 而得到的，更简便的形式则基于这两个模型的 R-平方。

8. 在计算 F 统计量时，分子 df 是被检验的约束个数，而分母 df 则是不受约束模型的自由度。

9. F 检验的对立假设是双侧对立假设。在经典方法中，我们规定一个显著性水平，它与分子 df 和分母 df 一起决定了临界值。当统计量 F 大于临界值 c 时，就拒绝虚拟假设。或者，我们可以计算一个 p 值，来概括拒绝 H_0 的证据。

10. 使用 F 统计量残差平方和的形式可以检验一般的多重线性约束。

154 11. 一个回归整体显著性的 F 统计量所检验的虚拟假设是，全部斜率参数都是零，而对截距则没有任何限制。在 H_0 下，解释变量对 y 的期望值没有任何影响。

关键术语

对立假设
经典线性模型
经典线性模型（CLM）假定
置信区间（CI）
临界值
分母自由度
经济显著性
排除性约束（限制）

分子自由度
单侧对立假设
单尾（侧）检验
p 值
实际重要性（显著性）
F 统计量的 R-平方型
拒绝法则
受约束模型

F 统计量
联合假设检验
联合不显著
联合统计显著
最小方差无偏估计量
多重假设检验
多重约束
正态性假定
虚拟假设
显著性水平
统计不显著
统计显著
t 比率
t 统计量
双尾（侧）对立假设
双尾检验
不受(无)约束模型

习 题

4.1 下面哪种因素可能导致通常 OLS 的 t 统计量无效（即在 H_0 下不服从 t 分布)?

(i) 异方差性。

(ii) 模型中两个自变量之间的样本相关系数达到 0.95。

(iii) 遗漏一个重要的解释变量。

4.2 考虑用企业年销售额、股本回报率（roe，以百分比表示）和企业股票的回报（ros，以百分比表示）来解释 CEO 薪水的一个方程：

$$\log(salary)=\beta_0+\beta_1\log(sales)+\beta_2 roe+\beta_3 ros+u$$

(i) 用模型的参数表述虚拟假设，即在控制了 $sales$ 和 roe 后，ros 对 CEO 的薪水没有影响。还表述对立假设：股票市场业绩更好，会提高 CEO 的薪水。

(ii) 使用 CEOSAL1.RAW 中的数据，通过 OLS 可以得到如下方程：

$$\begin{aligned}\log(\widehat{salary})&=\underset{(0.32)}{4.32}+\underset{(0.035)}{0.280}\log(sales)+\underset{(0.0041)}{0.0174}\,roe+\underset{(0.00054)}{0.00024}\,ros\\ n&=209,\ R^2=0.283\end{aligned}$$ 155

如果 ros 提高 50 点，预计 $salary$ 会提高多大比例？ros 对 $salary$ 具有实际上很大的影响吗？

(iii) 你最后会在一个用企业业绩表示 CEO 报酬的模型中包括 ros 吗？给出你的解释。

4.3 变量 $rdintens$ 是研发支出（R&D）占销售额的百分比；销售额以百万美元度量；变量 $profmarg$ 是利润占销售额的百分比。

利用 RDCHEM.RAW 中化工产业中 32 家企业的数据，估计如下方程：

$$\begin{aligned}\widehat{rdintens}&=\underset{(1.369)}{0.472}+\underset{(0.216)}{0.32}\log(sales)+\underset{(0.046)}{0.050}\,profmarg\\ n&=32,\ R^2=0.099\end{aligned}$$

(i) 解释 $\log(sales)$ 的系数。特别是，如果 $sales$ 增加 10%，估计 $rdintens$ 会变化多少个百分点？这在经济上是一个很大的影响吗？

(ii) 检验 R&D 强度不随 $sales$ 而变化，对立假设是，它随着销售额的增加而提高。在 5% 和 10% 的显著性水平上进行这个检验。

(iii) $profmarg$ 对 $rdintens$ 是否有统计显著的影响？

4.4 租金率是否受到一个大学城里学生人数的影响呢？令 $rent$ 表示美国一个大学城里单位租借面积的平均月租金；pop 表示总城市人口；$avginc$ 表示城市平均收入；$pctstu$ 表示学生人数占总人口的百分比。一个检验某种关系的模型是

$$\log(rent) = \beta_0 + \beta_1\log(pop) + \beta_2\log(avginc) + \beta_3 pctstu + u$$

(i) 表述虚拟假设：在其他条件不变的情况下，相对于总人口学生人数的多少对月租金没有影响。表述有影响的对立假设。

(ii) 你预期 β_1 和 β_2 具有什么样的符号？

(iii) 利用 RENTAL.RAW 中 64 个大学城在 1990 年的数据所估计的方程为

$$\begin{aligned}\widehat{\log(rent)} = {} & 0.043 + 0.066\log(pop)\\ & (0.844)\quad (0.039)\\ & + 0.507\log(avginc) + 0.0056\,pctstu\\ & (0.081)\qquad\qquad (0.0017)\\ & n = 64,\ R^2 = 0.458\end{aligned}$$

“总人口增加 10%将伴随以租金提高约 6.6%”这个说法有什么不妥？

156 (iv) 在 1% 的显著性水平上检验 (i) 部分陈述的假设。

4.5 考虑例 4.3 所估计的方程，它可以用来研究缺课对大学 GPA 的影响：

$$\begin{aligned}\widehat{colGPA} = {} & 1.39 + 0.412\,hsGPA + 0.015\,ACT - 0.083\,skipped\\ & (0.33)\quad (0.094)\qquad (0.011)\qquad (0.026)\\ & n = 141,\ R^2 = 0.234\end{aligned}$$

(i) 利用标准正态近似，求出（$hsGPA$ 在置信水平为 95% 时的置信区间。

(ii) 你能在 5% 的显著性水平上相对于双侧对立假设拒绝假设 H_0: $\beta_{hsGPA} = 0.4$ 吗？

(iii) 你能在 5% 的显著性水平上相对于双侧对立假设拒绝假设 H_0: $\beta_{hsGPA} = 1$ 吗？

4.6 在 4.5 节，我们使用了一个检验住房价格定价理性的例子。在那里，我们使用了 $price$ 和 $assess$ 的一个对数—对数型模型［参见方程 (4.47)］。这里，我们采用一个水平值—水平值的表述。

(i) 在简单回归模型

$$price = \beta_0 + \beta_1 assess + u$$

中，如果 $\beta_1 = 1$ 和 $\beta_0 = 0$，则评价是理性的。所估计的方程是

$$\widehat{price} = -14.47 + 0.976 assess$$
$$(16.27)\quad(0.049)$$
$$n = 88,\ \mathrm{SSR} = 165\ 644.51,\ R^2 = 0.820$$

首先，对双侧对立假设检验假设 H_0：$\beta_0 = 0$。然后，对双侧对立假设检验 H_0：$\beta_1 = 1$。你的结论是什么？

（ii）为了检验联合假设 $\beta_0 = 0$ 和 $\beta_1 = 1$，我们需要受约束模型的 SSR。这就要求在 $n = 88$ 的情况下计算 $\sum_{i=1}^{n}(price_i - assess_i)^2$，因为受约束模型的残差刚好就是 $price_i - assess_i$。（由于两个参数在 H_0 下都被设定，所以不需要受约束模型的估计值。）这最终得到 SSR = 209 448.99。对这个联合假设进行 F 检验。

（iii）现在检验模型

$$price = \beta_0 + \beta_1 assess + \beta_2 sqrft + \beta_3 lotsize + \beta_4 bdrms + u$$

的假设 H_0：$\beta_2 = 0$，$\beta_3 = 0$ 和 $\beta_4 = 0$。利用同样的 88 个住房数据估计这个模型的 R-平方是 0.829。

（iv）如果 *price* 的方程随着 *assess*，*sqrft*，*lotsize* 或 *bdrms* 而变化，你对第（iii）部分的 F 检验有什么看法？

157 4.7 在例 4.7 中，我们利用密歇根州制造业的数据估计了废弃率与其他企业特征之间的关系。我们现在来更深入地分析这个例子，并使用一个更大的企业样本。

（i）例 4.7 中待估计的总体模型可写成

$$\log(scrap) = \beta_0 + \beta_1 hrsemp + \beta_2 \log(sales) + \beta_3 \log(employ) + u$$

利用 1987 年的 43 个观测，所估计的方程是

$$\widehat{\log(scrap)} = 11.74 - 0.042 hrsemp - 0.951 \log(sales) + 0.992 \log(employ)$$
$$(4.57)\quad(0.019)\qquad(0.370)\qquad(0.360)$$
$$n = 43,\ R^2 = 0.310$$

（ii）证明这个总体模型也可以写成

$$\log(scrap) = \beta_0 + \beta_1 hrsemp + \beta_2 \log(sales/employ) + \theta_3 \log(employ) + u$$

式中，$\theta_3 \equiv \beta_2 + \beta_3$。［提示：$\log(x_2/x_3) = \log(x_2) - \log(x_3)$。］解释假设 H_0：$\theta_3 = 0$。

（iii）当估计第（ii）部分的方程时，我们得到

$$\widehat{\log(scrap)} = 11.74 - 0.042 hrsemp$$
$$(4.57)\quad(0.019)$$
$$- 0.951 \log(sales/employ) + 0.041 \log(employ)$$

$$\begin{array}{lr}(0.370) & (0.205)\\ \multicolumn{2}{l}{n=43,\ R^2=0.310}\end{array}$$

控制了工人培训和销售—雇员比后，是否越大的企业具有越为统计显著的废弃率？

(iv) 检验假设：*sales/employ* 提高1%将伴随以废弃率下降1%。

4.8 在经典线性模型假定 MLR.1～MLR.6 下，考虑含有三个自变量的多元回归模型：

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+u$$

你想检验的虚拟假设是 H_0：$\beta_1-3\beta_2=1$。

(i) 令 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 表示 β_1 和 β_2 的OLS估计量。用 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差及其协方差求出 $\mathrm{Var}(\hat{\beta}_1-3\hat{\beta}_2)$。$\hat{\beta}_1-3\hat{\beta}_2$ 的标准误是什么？

(ii) 写出检验 H_0：$\beta_1-3\beta_2=1$ 的 t 统计量。

(iii) 定义 $\theta_1=\beta_1-3\beta_2$ 和 $\hat{\theta}_1=\hat{\beta}_1-3\hat{\beta}_2$。写出一个涉及 β_0，θ_1，β_2 和 β_3 的回归方程，使你能直接得到 $\hat{\theta}_1$ 及其标准误。

4.9 在习题3.3中，我们估计了方程

$$\begin{array}{llll}\widehat{sleep}=3\,638.25-0.148\,totwrk & -11.13\,educ & +2.20\,age\\ \quad(112.28)\quad(0.017) & \quad(5.88) & \quad(1.45)\\ \quad n=706,\ R^2=0.113\end{array}$$

158 式中的标准误是我们现在才同估计值一并给出的。

(i)相对于一个双侧对立假设，是 *educ* 还是 *age* 在5%的水平上是个别显著的？给出你的计算。

(ii)从方程中去掉 *educ* 和 *age*，则给出

$$\begin{array}{l}\widehat{sleep}=3\,586.38-0.151\,totwrk\\ \quad(38.91)\quad(0.017)\\ \quad n=706,\ R^2=0.103\end{array}$$

在5%的显著性水平上，*educ* 和 *age* 在原方程中是联合显著的吗？说明你所给答案的理由。

(iii) 在模型中包括 *educ* 和 *age*，是否显著影响所估计的睡眠和工作之间的替换关系？

(iv) 假设睡眠方程含有异方差性。这对第(i)和(ii)部分计算的检验意味着什么？

4.10 回归分析还可以用来检验市场是否在评价股票时有效地使用了市场信息。为简单起见，令 *return* 为持有一个企业的股票在从1990年末到1994年末的四年时间内得到的总回报。**有效市场假设**认为，这些回报不应该与1990年知道的信息存在系统相关。如果在期初知道的企业特征有助于预测股票回报，那我们在选择股票时就能用到这个信息。

对于1990年，令 *dkr* 表示企业的债务—资本比率；*eps* 表示每股收益；(log) *netinc* 表示净收入；而(log) *salary* 则表示CEO的总报酬。

（i）使用 RETURN.RAW 中的数据，估计了如下方程：

$$\widehat{return} = 40.44 + 0.952dkr + 0.472eps - 0.025netinc + 0.003salary$$
$$(29.30)\quad(0.854)\quad(0.332)\quad(0.020)\quad(0.009)$$
$$n = 142,\ R^2 = 0.028\,5$$

检验这些解释变量在 5% 的显著性水平上是否联合显著。哪个解释变量是个别显著的？

（ii）现在使用 *netinc* 和 *salary* 的对数形式重新估计这个模型：

$$\widehat{return} = -69.12 + 1.056dkr + 0.586eps - 31.18netinc + 39.26salary$$
$$(164.66)\quad(0.847)\quad(0.336)\quad(14.16)\quad(26.40)$$
$$n = 142,\ R^2 = 0.053\,1$$

第（i）部分的结论有没有什么变化？

（iii）总的看来，股票回报可预测性的证据是强还是弱？

4.11　使用 CEOSAL2.RAW 中的数据得出下表：

159

	因变量：log（*salary*）		
自变量	(1)	(2)	(3)
log（*sales*）	0.224	0.158	0.188
	(0.027)	(0.040)	(0.040)
log（*mktval*）	—	0.112	0.100
		(0.050)	(0.049)
profmarg	—	−0.002 3	−0.002 2
		(0.002 2)	(0.002 1)
ceoten	—	—	0.017 1
			(0.005 5)
comten	—	—	−0.009 2
			(0.003 3)
intercept	4.94	4.62	4.57
	(0.20)	(0.25)	(0.25)
观测个数	177	177	177
R-平方	0.281	0.304	0.353

变量 *mktval* 为企业的市场价值；*profmarg* 为利润占销售额的百分比；*ceoten* 为其就任当前公司 CEO 的年数；而 *comten* 则为其在这个公司的总年数。

（i）评论 *profmarg* 对 CEO 薪水的影响。

（ii）市场价值是否具有显著影响？试解释你的结论。

（iii）解释 *ceoten* 和 *comten* 的系数。这些变量是统计显著的吗？你如何解释在其他条件不变的情况下，你在这个公司任职时间越长，你的薪水则越低？

计算机习题

4.12　如下模型可用来研究竞选支出如何影响选举结果：

$$voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u$$

式中，$voteA$ 为候选人 A 得到的选票百分数；$expendA$ 和 $expendB$ 分别为候选人 A 和 B 的竞选支出；而 $prtystrA$ 则为对 A 所在党派之实力的一种度量
160 （A 所在党派在最近一次总统选举中获得的选票百分比）。

（i）如何解释 β_1？

（ii）用参数表述如下虚拟假设：A 的竞选支出提高 1% 被 B 的支出提高 1% 所抵消。

（iii）利用 VOTE1.RAW 中的数据来估计上述模型，并以通常的方式报告结论。A 的竞选支出会影响结果吗？B 的支出呢？你能用这些结论来检验第（ii）部分中的假设吗？

（iv）估计一个模型，使之能直接给出检验第（ii）部分中假设所需用的 t 统计量。你有什么结论？（使用双侧对立假设。）

4.13　本题要利用 LAWSCH85.RAW 中的数据。

（i）使用与习题 3.4 一样的模型，表述并检验虚拟假设：在其他条件不变的情况下，法学院排名对起薪的中位数没有影响。

（ii）新生年级的学生特征（即 $LSAT$ 和 GPA），对解释 $salary$ 而言是个别或联合显著的吗？

（iii）检验是否要在方程中引入入学年级的规模（$clsize$）和教职工的规模（$faculty$）；只进行一个检验。（注意解释 $clsize$ 和 $faculty$ 方面的数据缺失。）

（iv）还有哪些因素可能影响到法学院的排名，但又没有包括在薪水回归中？

4.14　参考习题 3.14。现在，我们使用住房价格的对数作为因变量：

$$\log(price) = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

（i）你想在住宅中增加一间 150 平方英尺的卧室的情况下，估计并得到 $price$ 变化百分比的一个置信区间；以小数形式表示就是 $\theta_1 = 150\beta_1 + \beta_2$。使用 HPRICE1.RAW 中的数据去估计 θ_1。

（ii）用 θ_1 和 β_1 表达 β_2，并代入 log（$price$）的方程。

（iii）利用第（ii）部分中的结果，得到 $\hat{\theta}_1$ 的标准误，并使用这个标准误构造一个 95% 的置信区间。

4.15　在例 4.9 中，可以使用样本中所有的 1 388 个观测数据去估计那个受约束模型。使用所有观测值，计算 $bwght$ 对 $cigs$，$parity$ 和 $faminc$ 回归的 R-平方，并与例 4.9 中受约束模型所报告的 R-平方相比较。

4.16　本题要用到 MLB1.RAW 中的数据。

（i）使用方程（4.31）中所估计的模型，并去掉变量 $rbisyr$，$hrunsyr$ 的统计显著性会怎么样？$hrunsyr$ 的系数大小又会怎么样？

（ii）在第（i）部分的模型中增加变量 *runsyr*，*fldperc* 和 *sbasesyr*；这些因素中的哪一个是显著的？

（iii）在第（ii）部分的模型中，检验 *bavg*，*fldperc* 和 *sbasesyr* 的联合显著性。

161 4.17 本题要用到 WAGE2.RAW 中的数据。

（i）考虑标准的工资方程

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 tenure+u$$

表述如下虚拟假设：多一年一般性的工作经历，与为现任雇主多工作一年相比，对$\log(wage)$的影响是相同的。

（ii）通过构造一个95%的置信区间，在5%的显著性水平上，相对于双侧对立假设检验第（i）部分中的虚拟假设。你的结论是什么？

第5章　多元回归分析：OLS的渐近性

162　在第3章和第4章中，我们讨论了总体模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{5.1}$$

中OLS估计量的所谓有限样本或小样本的精确性质。比如，OLS在前四个高斯-马尔科夫假定下的无偏性就是一个有限样本的性质，因为它对任何样本容量n（略加限制的地方是，n必须至少和回归模型中参数的总个数$k+1$一样大）都成立。类似地，OLS在全套高斯-马尔科夫假定（MLR.1～MLR.5）下是最优线性无偏估计量的事实，也是一个有限样本性质。

在第4章，增加了经典线性模型假定MLR.6，它表明误差项u服从正态分布并独立于解释变量。这就使我们能够（以样本中的解释变量为条件）推导出OLS估计量精确的抽样分布。特别是，定理4.1表明OLS估计量具有正态的抽样性质，这就直接产生了t统计量和F统计量的t分布和F分布。如果误差不是正态分布的，那么，对任意的样本容量，t统计量的分布就不再恰好是t分布，F统计量也不再完全服从一个F分布。

除了有限样本性质外，了解估计量和检验统计量的**渐近性质**（asymptotic properties）或**大样本性质**（large sample properties）也很重要。定义这些性质不是针对特定的样本容量，而是针对样本容量无限增加的情况。幸运的是，在我们所做的假定下，OLS具有令人满意的大样本性质。一个实践中相

当重要的发现是，即使没有正态性假定（假定 MLR.6），t 统计量和 F 统计量也近似服从 t 分布和 F 分布，至少在大样本容量的情况下如此。我们在 5.1 节讨论过 OLS 的一致性问题之后，会在 5.2 节更详细地讨论这一点。

5.1 一致性

估计量的无偏性固然重要，但并非总能实现。比如，像我们在第 3 章讨论过的那样，在多元回归模型中，回归标准误 $\hat{\sigma}$ 就不是误差 u 的标准差 σ 的一个无偏估计量。尽管 OLS 估计量在假定 MLR.1～MLR.4 下是无偏的，但我们在第 11 章将发现，OLS 估计量在时间序列回归中会失去无偏性。而且，
163 在本书的第 3 部分，我们还会遇到其他几个有偏误的估计量。

既然并非所有有用的估计量都是无偏的，所以几乎所有的经济学家都同意**一致性**（consistency）是对一个估计量的最起码要求。著名计量经济学家克里夫·W·J·葛兰杰（Clive W.J.Granger）曾说过："如果你在 n 趋于无穷时还不能正确地得到它，那你就不应该做这件事。"他的意思是说，如果你给出一特定总体参数的估计量不是一致的，那你是在浪费时间。

描述一致性有几种不同的方法。规范的定义和结论都在附录 C 中给出，我们在这里只强调直觉上的理解。为简洁起见，对某个 j，令 $\hat{\beta}_j$ 表示 β_j 的 OLS 估计量。对每个 n，$\hat{\beta}_j$ 都有一个概率分布（代表其在不同随机样本容量下的可能值）。由于 $\hat{\beta}_j$ 在假定 MLR.1～MLR.4 下是无偏的，所以这个分布的均值就是 $\hat{\beta}_j$。如果估计量是一致的，那么随着样本容量的增加，$\hat{\beta}_j$ 的分布就越来越紧密地分布在 β_j 的周围。当 n 趋于无穷时，$\hat{\beta}_j$ 的分布就紧缩成单一个点 β_j。实际上，这意味着，如果能搜集到我们所需要的样本数，就能让我们的估计量任意接近于 β_j。图 5.1 说明了这种收敛性。

164 对于任何一个具体的应用而言，我们都自然而然地有一个固定的样本容量，这就是难以理解诸如一致性等渐近性质的原因。一致性涉及的一个设想的实验，考虑的是样本容量变大时发生的情况（而另一方面我们得到的是每个样本容量的大量随机样本）。如果得到越来越多的数据，还不能让我们更接近我们所关心的估计值，那就是因为我们所使用的估计程序表现欠佳。

恰巧，同样的那一系列假定同时蕴涵着 OLS 的无偏性和一致性。让我们用一个定理加以总结。

定理 5.1（OLS 的一致性）

在假定 MLR.1～MLR.4 下，对所有的 $j=0, 1, \cdots, k$，OLS 估计量 $\hat{\beta}_j$ 都是 β_j 的一致估计。

利用附录 D 和 E 中所介绍的矩阵代数，很容易就能给出上述结论的一般性证明，但我们在简单回归模型中可以毫不费力地证明定理 5.1。我们主要关注的是斜率估计量 $\hat{\beta}$。

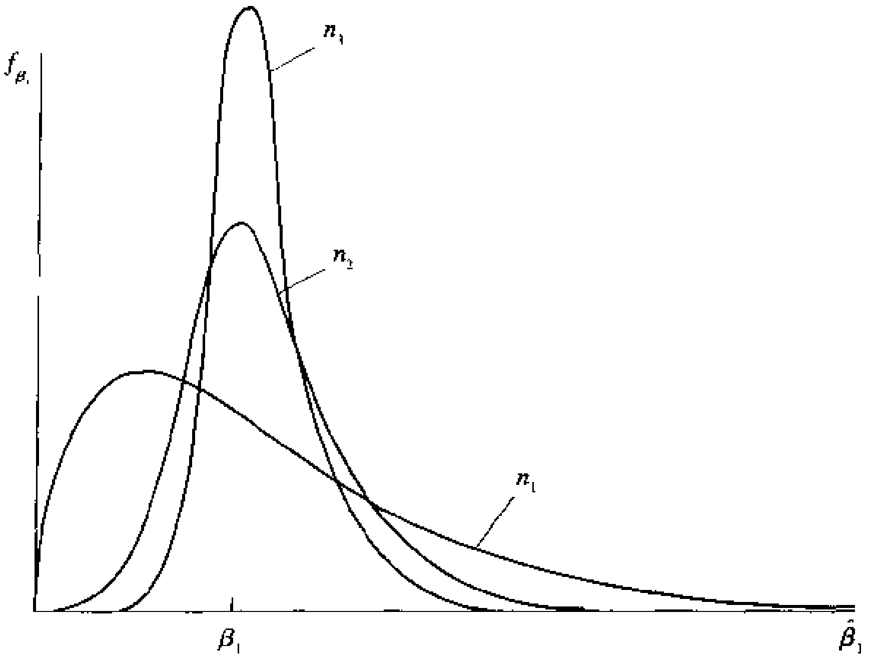


图 5.1　$\hat{\beta}_1$ 在样本容量 $n_1<n_2<n_3$ 情况下的抽样分布

证明过程与对无偏性的证明相同：写下 $\hat{\beta}_1$ 的公式，然后将 $y_i=\beta_0+\beta_1x_{i1}+u_i$ 代入，得到

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}$$

$$=\beta_1+\frac{n^{-1}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)u_i}{n^{-1}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2} \tag{5.2}$$

我们可以在分子和分母中应用大数定律，则分别在概率上收敛于总体值 $\mathrm{Cov}(x_1,u)$ 和 $\mathrm{Var}(x_1)$。给定 $\mathrm{Var}(x_1)\neq0$（MLR.4 对此做了假定），我们可以使用概率极限的性质（参见附录 C），得到

$$\mathrm{plim}\,\hat{\beta}_1=\beta_1+\mathrm{Cov}(x_1,u)/\mathrm{Var}(x_1)=\beta_1\text{，因为 }\mathrm{Cov}(x_1,u)=0 \tag{5.3}$$

我们已经用到了第 2 章和第 3 章所讨论的结论，即 $\mathrm{E}(u|x_1)$ 意味着 x_1 和 u 不相关（协方差为零）。

作为一个技术问题，为保证概率极限的存在，还应该假定 $\mathrm{Var}(x_1)<\infty$ 和 $\mathrm{Var}(u)<\infty$（即它们的概率分布不至于过于分散），但我们不去担心这些假定可能不成立的情况。

前面的论证，特别是方程（5.3）表明，如果我们假定只有零相关，那么 OLS 在简单回归情形中就是一致的。在一般情形中也是这样，我们现在把这一点表述成一个假定。

假定 MLR.3′（零均值和零相关）

对所有的 $j=1, 2, \cdots, k$，都有 $\mathrm{E}(u)=0$ 和 $\mathrm{Cov}(x_j, u)=0$。

165 我们在第 3 章讨论了为什么假定 MLR.3 意味着 MLR.3′，但反过来则不成立。OLS 在更弱的假定 MLR.3′下仍是一致估计的事实，最终在第 15 章和其他情形中将有用武之地。有趣的是，尽管 OLS 在假定 MLR.3 下是无偏的，但在假定 MLR.3′下则不然。（这正是我们过去假定 MLR.3 的主要原因。）

推导 OLS 的不一致性

就像 $\mathrm{E}(u \mid x_1, x_2, \cdots, x_k)=0$ 不成立能导致 OLS 估计量的偏误一样，u 和 $x_1, x_2, \cdots x_k$ 中的任何一个相关，通常也能导致所有的 OLS 估计量都失去其一致性。通常可以把这个简单而又重要的观察总结为：**如果误差与任何一个自变量相关，那么 OLS 就是有偏而又不一致的估计**。这是极为不幸的，因为它就意味着，随着样本容量的增大，偏误将继续存在。

在简单回归情形中，我们可以从方程（5.3）（无论 u 和 x_1 相关与否，这个方程的第一个等号都成立）中得到不一致性。$\hat{\beta}_1$ 的**不一致性**（inconsistency）［有时也粗略地称为**渐近偏误**（asymptotic bias）］为

$$\operatorname{plim}\hat{\beta}_1-\beta_1=\frac{\mathrm{Cov}(x_1, u)}{\mathrm{Var}(x_1)} \tag{5.4}$$

因为 $\mathrm{Var}(x_1)>0$，所以，若 x_1 和 u 正相关，则 $\hat{\beta}_1$ 的不一致性就为正，而若 x_1 和 u 负相关，则 $\hat{\beta}_1$ 的不一致就为负。如果 x_1 和 u 之间的协方差相对于 x_1 的方差很小，那么这种不一致就可以被忽略；不幸的是，由于 u 是观测不到的，所以我们甚至还不能估计出这个协方差有多大。

我们可以使用式（5.4）来推导遗漏变量偏误（参见第 3 章的表 3.2）的渐近类似情况。假设真实模型

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+v$$

满足前四个高斯-马尔科夫假定，那么 v 的均值就是零，而且与 x_1 和 x_2 都不相关。如果 $\hat{\beta}_0$，$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 表示 y 对 x_1 和 x_2 回归所得到的 OLS 估计量，那么定理 5.1 则意味着这些估计量都是一致的。如果我们在回归中漏掉 x_2 而将 y 对 x_1 进行简单回归，那么 $u=\beta_2 x_2+v$。令 $\widetilde{\beta}_1$ 表示简单回归的斜率估计量。于是

$$\operatorname{plim}\widetilde{\beta}_1=\beta_1+\beta_2\delta_1 \tag{5.5}$$

式中

$$\delta_1=\frac{\mathrm{Cov}(x_1, x_2)}{\mathrm{Var}(x_1)} \tag{5.6}$$

因此，我们实际上可以把这种不一致性看成偏误。区别在于，不一致是用

x_1 的总体方差与 x_1 和 x_2 之间的总体协方差表示的，而偏误则基于其样本对应量（因为我们以样本中 x_1 和 x_2 的值为条件）。

166 如果 x_1 和 x_2（在总体中）不相关，那么 $\delta_1=0$，从而 $\tilde{\beta}_1$ 是 β_1 的一致估计量（尽管不一定是无偏的）。如果 x_2 对 y 具有正的偏效应，以致 $\beta_2>0$；而且 x_1 和 x_2 正相关，以致 $\delta_1>0$，那么，$\tilde{\beta}_1$ 的不一致性便是正偏的，如此等等。我们可以从表 3.2 得到不一致性或渐近偏误的方向。如果 x_1 和 x_2 之间的协方差相对 x_1 的方差而言很小，那么这种不一致性也会很小。

例 5.1　住房价格与住房到垃圾焚化炉的距离

令 y 表示一套住房的价格（*price*），x_1 表示该住房到一座新的垃圾焚化炉的距离（*distance*），而 x_2 表示住房的“质量”（*quality*）。变量 *quality* 有些含糊，可包括住房和客厅的大小、卧室和浴室的间数，以及诸如邻里关系的吸引力等无形价值。如果焚化炉使住房价格下降，那么 β_1 就应该为正：所有其他情况不变，远离焚化炉的住房更值钱。按定义，由于在其他因素不变的情况下，质量越高的住房卖得越贵，所以 β_2 为正。如果焚化炉基本上都远离较好的住家而建，那么 *distance* 和 *quality* 就正相关，所以 $\delta_1>0$。*price* 对 *distance* [或 log（*price*）对log（*distance*）]的简单回归倾向于高估焚化炉的影响：$\beta_1+\beta_2\delta_1>\beta_1$。

对 OLS 估计量来说，重要的一点是，根据定义，这个问题不会随着在样本中增加更多的观测而消失。更多的数据有可能使这个问题变得更糟：随着样本容量的增加，OLS 估计量会越来越接近于 $\beta_1+\beta_2\delta_1$。

问题 5.1

假设模型

$$score=\beta_0+\beta_1 skipped+\beta_2 priGPA+u$$

满足前四个高斯-马尔科夫假定，其中，*score* 为期末考试分数；*skipped* 为逃课次数；*priGPA* 为上一学期的 GPA。如果 $\tilde{\beta}_1$ 是 *score* 对 *skipped* 的简单回归所得到的估计量，那么其渐近偏误的方向是什么？

在含有 k 个回归元的情形中，因为推导偏误变得很困难，对不一致性的偏向和大小的推测也要困难得多。我们要记住，如果我们有方程（5.1）中的模型，其中，比方说，x_1 和 u 相关而其他自变量和 u 不相关，那么所有的 OLS 估计量一般地说都是不一致的。比如，在 $k=2$ 的情形中，有

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u$$

假设 x_2 和 u 不相关，但 x_1 和 u 相关。于是，OLS 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 一般都是不一致的。（这个截距的估计量也将是不一致的。）x_1 和 x_2 通常是相关的，从而引起 $\hat{\beta}_2$ 的不一致性。如果 x_1 和 x_2 不相关，那么 x_1 和 u 之间的任意相关都不会导致 $\hat{\beta}_2$ 的不一致性：$\mathrm{plim}\hat{\beta}_2=\beta_2$。此外，$\hat{\beta}_1$ 的不一致性和式（5.4）中的一样。在一般情形中，同理可推知，如果 x_1 与 u 相关，但 x_1

和 u 与其他自变量都不相关，那么就只有 $\hat{\beta}_1$ 是不一致的，而且这个不一致性仍由式（5.4）给出。

5.2 渐近正态和大样本推断

167 虽然估计量的一致性是一个重要性质，但仅有一致性还不足以使我们进行统计推断。仅知道估计量随着样本容量的扩大越来越接近总体值，还不足以让我们对参数假设进行检验。为了进行检验，我们还需要 OLS 估计量的抽样分布。在经典线性模型假定 MLR.1～MLR.6 下，定理 4.1 表明，抽样分布是正态的。这个结论，正是推导我们在应用计量经济学中常用的 t 分布和 F 分布的基础。

OLS 估计量确切的正态性关键取决于总体中误差（u）分布的正态性。如果误差 u_1，u_2，…，u_n 是从某个非正态分布中随机抽取的，那么 $\hat{\beta}_j$ 也不会正态地分布，这就意味着，t 统计量不具有 t 分布，而且 F 统计量不具有 F 分布。由于我们的推断取决于我们能否从 t 分布或 F 分布中得到临界值或 p 值，所以这是一个潜在的严重问题。

记得假定 MLR.6 等于是说，给定 x_1，x_2，…，x_k，y 的分布是正态的。因为在一个特定的应用中可以观测到 y 而观测不到 u，所以考虑 y 的分布有没有可能是正态的则容易得多。事实上，我们已经看到了几个 y 绝不可能具有正态分布的例子。一个正态分布的随机变量，对称地分布在其均值两边，它可以取任何一个正值或负值（但取每个值的概率都是零），而且分布中超过 95% 的面积处在 2 倍的标准差之内。

在例 3.4 中，我们估计了一个模型，用来解释青年人在一特定年份被拘捕的次数（*narr*86）。在总体中，多数人在这一年没有被拘捕过，而且绝大多数人最多只被拘捕过一次。（在数据集 CRIME1.RAW 中由 2 725 个人构成的样本中，只有不到 8% 的人在 1986 年被拘捕过一次以上。）由于对样本中 92% 的人来说，变量 *narr*86 都只取两个值，所以它在总体中不可能接近正态分布。

在例 4.6 中，我们估计了另一个模型，用来解释 401k 养老金计划的参与率（*prate*）。图 5.2 中的频率分布（也被称为直方图）表明，*prate* 的分布明显向右偏斜，而不是正态分布。事实上，在对 *prate* 的观测中，有超过 40% 都取值 100，即 100% 地参与。即便以解释变量为条件，这也违背正态性假定。

168 我们知道，正态性对 OLS 的无偏性不起任何作用，也不影响 OLS 在高斯-马尔科夫假定下是最优线性无偏估计的结论。但基于 t 统计量或 F 统计量的准确推断还要求 MLR.6。这是否意味着，在例 4.6 对 *prate* 的分析中，我们在决定哪个变量统计显著时必须放弃 t 统计量呢？幸运的是，对这个问题的回答是否定的。尽管 y_i 不是来自一个正态分布，但我们可以利用附录

C 中的中心极限定理断定，至少在大样本容量的情况下，OLS 估计量是近似正态分布的。

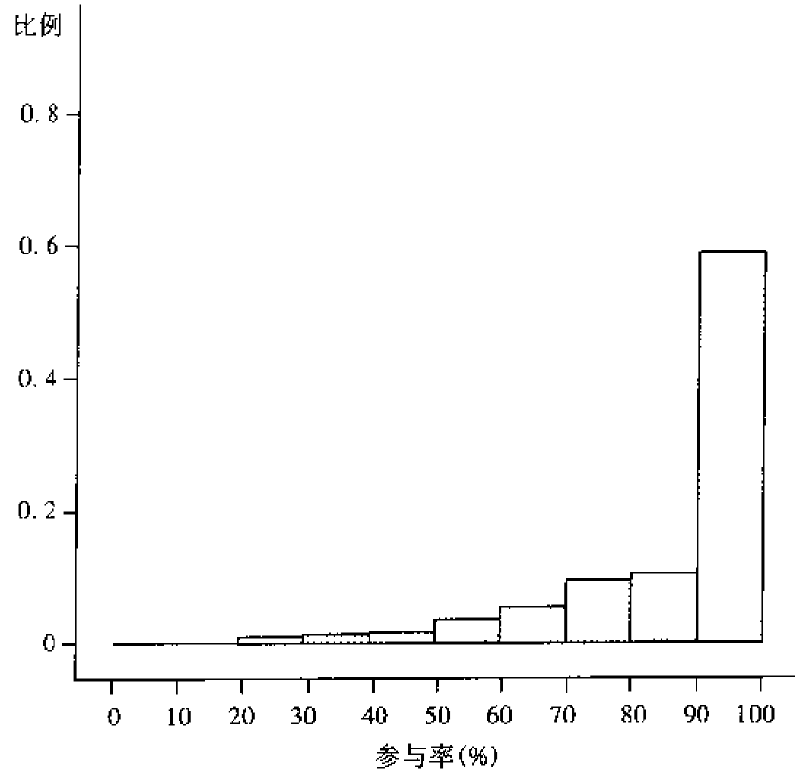


图 5.2 使用 401K.RAW 中的数据得到的 *prate* 的直方图

定理 5.2（OLS 的渐近正态性）

在高斯-马尔科夫假定 MLR.1～MLR.5 下：

(i) $\sqrt{n}(\hat{\beta}_j-\beta_j)\overset{a}{\sim}\text{Normal}(0,\sigma^2/a_j^2)$，其中 $\sigma^2/a_j^2>0$ 是 $\sqrt{n}(\hat{\beta}_j-\beta_j)$ 的渐近方差；至于斜率系数，$a_j^2=\text{plim}\left(n^{-1}\sum_{i=1}^{n}\hat{r}_{ij}^2\right)$，其中 $\hat{r}_{ij}$ 是 x_j 对其余自变量进行回归所得到的残差。我们便说 $\hat{\beta}_j$ 是渐近正态分布的（参见附录 C）。

(ii) $\hat{\sigma}_2$ 是 $\hat{\sigma}=\text{Var}(u)$ 的一个一致估计量；

(iii) 对每个 j，有

$$(\hat{\beta}_j-\beta_j)/\text{se}(\hat{\beta}_j)\overset{a}{\sim}\text{Normal}(0,1) \tag{5.7}$$

其中 $\text{se}(\hat{\beta}_j)$ 就是通常的标准误。

169 对渐近正态的证明多少有些复杂，在附录中也只对简单回归情形勾勒了一下证明过程。第（ii）部分得自于大数定律，第（iii）部分则得自于第（i）、第（ii）部分和附录 C 中所讨论的渐近性质。

定理 5.2 的重要之处在于，它去掉了正态性假定 MLR.6；对误差分布惟一的限制是，它具有有限方差，这一点我们将一直这样假定。我们还对 u

假定了零条件均值和同方差性。

注意到标准正态分布在式（5.7）中出现的方式与 t_{n-k-1} 分布不同。这是因为，这个分布只是一个近似。对比之下，在定理4.2中，式（5.7）中的比率的分布对于任何样本容量都恰好是 t_{n-k-1}。从实践的角度来看，这个差别并不重要。实际上，由于随着自由度的变大，t_{n-k-1}会趋近于标准正态分布，所以如下写法也是合理的：

$$(\hat{\beta}_j - \beta_j)/\mathrm{se}(\hat{\beta}_j) \overset{a}{\sim} t_{n-k-1} \tag{5.8}$$

方程（5.8）告诉我们，进行 t 检验和构造置信区间，都与在经典线性模型的假定下完全一样进行。这意味着，如果高斯-马尔科夫假定成立，我们对像 *prate* 和 *narr*86 之类因变量的分析，根本就没有必要改变：在这两种情况下，我们都至少有1 500个观测，这肯定足以证明中心极限定理的近似是正确的。

如果样本容量不是很大，那么，当 u 不是正态分布时，t 分布可能是 t 统计量分布的一个不好的近似。不幸的是，在近似程度足够好之前，对于样本容量必须有多大，并没有一般性的规定。有些计量经济学家认为 $n=30$ 就令人满意了。但这不可能足以对付 u 所有可能的分布。根据 u 的分布，在应用中心极限定理之前，多一些观测可能是必要的。而且，近似的质量不仅取决于 n，还取决于自由度 $n-k-1$：模型中的自变量越多，使用近似通常就需要越大的样本容量。在自由度很小和非正态误差情况下的推断方法超出了本书的范围。我们将像平常那样，干脆使用 t 统计量，而不去担心正态性假定。

看出定理5.2必须要求同方差性假定（与零条件均值假定一起）这一点很重要。如果 $\mathrm{Var}(y|x)$ 不是常数，那么无论样本容量有多大，通常的 t 统计量和置信区间都是无效的；在出现异方差性时，中心极限定理也不能使我们摆脱困境。基于这个原因，我们将用整个第8章来讨论在出现异方差性时的处理办法。

定理5.2的结论之一是，$\hat{\sigma}^2$ 是 σ^2 的一个一致估计量；我们从定理3.3已经知道，在高斯-马尔科夫假定下，$\hat{\sigma}^2$ 是 σ^2 无偏估计量。一致性意味着，$\hat{\sigma}$ 是 σ 的一个一致估计量，这对构建方程（5.7）中渐近正态的结论而言是很重要的。

170 记住 $\hat{\sigma}$ 出现在每个 $\hat{\beta}_j$ 的标准误中。实际上，$\hat{\beta}_j$ 的估计方差是

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}_2}{\mathrm{SST}_j(1-R_j^2)} \tag{5.9}$$

式中，SST_j 为样本中 x_j 的总平方和，而 R_j^2 则是将 x_j 对其余所有自变量进行回归所得到的 R-平方。在3.4节，我们研究了式（5.9）中的每个组成部分，现在将在渐近分析的背景下进行阐述。随着样本容量的扩大，$\hat{\sigma}^2$ 概率收敛于常数 σ^2。而且，R_j^2 也趋近于一个严格介于0～1之间的数（所以 $1-R_j^2$ 也收敛于0～1之间的某个数）。x_j 的样本方差是 SST_j/n，所以 $\mathrm{SST}_j/$

n 也随着样本容量的扩大而收敛于 Var（x_j）。这意味着，SST_j 增加的速度近似于样本容量增加的速度：$\mathrm{SST}_j \approx n\sigma_j^2$，其中 σ_j^2 是 x_j 的总体方差。综合上述结论我们发现，$\mathrm{Var}(\hat{\beta}_j)$以速度 $1/n$ 收缩于零；这就说明了为什么大样本要好些。

问题 5.2 在一个样本容量很大的回归模型中，在假定 MLR.1～MLR.5 下，$\hat{\beta}_j$ 95% 水平上的近似置信区间是什么？我们把这个置信区间称为**渐近置信区间**（asymptotic confidence interval）。

当 u 不是正态分布时，式（5.9）的平方根有时被称为**渐近标准误**（asymptotic standard error），而 t 统计量也被称为**渐近 t 统计量**（asymptotic t statistics），因为它们和我们在第 4 章讨论的两个量一样，所以我们就把它们称为标准误和 t 统计量，只是在理解时，知道它们有时只在大样本情况下才是正确的。

使用前面有关估计方差的论述，我们可以写出

$$\mathrm{se}(\hat{\beta}_j) \approx \frac{c_j}{\sqrt{n}} \tag{5.10}$$

式中，c_j 是一个不依赖于样本容量的常数，且为正。虽然方程（5.10）只是一个近似，但它是一个有用的经验法则：可以预期标准误的收缩速度为样本容量平方根的倒数。

例 5.2　婴儿出生体重方程中的标准误

我们利用 BWGHT.RAW 中的数据估计了一种关系，其中因变量是婴儿出生时体重的对数，自变量是每天吸烟的数量（*cigs*）和家庭收入的对数（*faminc*）。观测总数为 1 388 个。使用前面一半的观测（694 个）得到 $\hat{\beta}_{cigs}$ 的标准误约为 0.001 3。使用全部观测所得到的标准误约为 0.000 86。后面一个标准误与前一个标准误的比为 0.000 86/0.001 3≈0.662。这相当接近于从式（5.10）中近似得到的比率 $\sqrt{694/1388} \approx 0.707$。换句话说，方程（5.10）意味着使用较大样本容量的标准误，应该约为使用较小样本容量之标准误的 70.7%。这个百分比相当接近于我们从标准误的比率中计算而来的比率 66.2%。

171 OLS 估计量的渐近正态还意味着，在大样本容量情况下，F 统计量具有近似的 F 分布。因此，对排除性约束和其他多元假设的检验，与我们前面的做法别无二致。

其他大样本检验：拉格朗日乘数统计量

我们一旦步入渐近分析的殿堂，就还有其他的检验统计量可用于假设检验。多数情况下，没有理由去用 t 和 F 之外的统计量：如我们所见，这些统计量无须正态性假定而具有大样本下的正确性。不过有时候用其他方法检验多元排除约束也很有用，我们现在就来讨论**拉格朗日乘数 *LM* 统计量**（Lagrange multiplier *LM* statistic），它在现代计量经济学中已经受到一定程度的欢迎。

“拉格朗日乘数统计量”一词源于约束条件下的最优化问题，这个专题超出了本书的研究范围。［参见 Davidson and MacKinnon（1993）。］有时也用**得分统计量**（score statistic）的说法（也是来自使用微积分的最优化理论）。幸运的是，在线性回归的框架下，无须翻阅复杂的数学就能很简单地推导 *LM* 统计量。

我们这里推导的 *LM* 统计量的形式依赖于高斯-马尔科夫假定，也就是那些在大样本条件下使得 F 统计量有效的假定。我们不需要正态性假定。

为了推导 *LM* 统计量，考虑通常包含 k 个自变量的多元回归模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{5.11}$$

我们想检验的是，这些变量中最后的 q 个是否在总体中其参数都是零。虚拟假设是

$$H_0: \beta_{k-q+1} = 0, \cdots, \beta_k = 0 \tag{5.12}$$

它对模型（5.11）施加了 q 个排除性约束。就像 F 检验一样，式（5.12）的对立假设是，这些参数中至少有一个异于零。

LM 统计量仅要求对受约束模型的估计。于是，假定我们进行了如下回归

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_{k-q} x_{k-q} + \tilde{u} \tag{5.13}$$

式中，“～”表示估计值都来自受约束模型。特别是 $\tilde{u}$ 表示了受约束模型的残差。（和往常一样，这无非表示我们对样本中的每个观测所得到的约束残差。）

如果被省略的变量 x_{k-q-1} 到 x_k 在总体中的系数都为零，那么 $\tilde{u}$ 应该与样本中这些变量中的每一个都不相关，至少近似于不相关。这就建议我们将这些残差对那些在 H_0 下被排除的变量进行一个回归，*LM* 检验基本上就是这样做的。然而结果表明，为了得到一个能使用的检验统计量，我们必须
172 在回归中包括所有的自变量（这样做的原因是技术性的，而且也不重要）。于是我们进行

$$\tilde{u} \text{ 对 } x_1, x_2, \cdots, x_k \text{ 的回归} \tag{5.14}$$

这就是**辅助回归**（auxiliary regression）的一个例子，辅助回归是用来计算一个检验统计量，但回归系数没有直接的意义。

我们如何使用式（5.14）中的回归结果来检验式（5.12）呢？如果式（5.12）是正确的，那么由于 $\tilde{u}$ 近似地与所有自变量都不相关，所以在存在抽样误差的情况下，式（5.14）的 R-平方也应该“接近”于零。与通常的假设检验一样，问题是，如何决定在一个选定的显著性水平下，统计量大到什么程度才足以拒绝虚拟假设。结果是，样本容量乘上辅助回归式（5.14）的 R-平方，渐近服从一个自由度为 q 的 chi 平方随机变量的分布。这就给出了检验一组 q 个自变量联合显著性的简单程序。

q 个排除性约束的拉格朗日乘数统计量

(i) 将 y 对施加限制后的自变量进行回归，并保存残差 $\tilde{u}$。

(ii) 将 $\tilde{u}$ 对所有自变量进行回归，并得到 R-平方，记为 R_u^2（以区别于将 y 作为因变量时所得到的 R-平方）。

(iii) 计算 $LM = nR_u^2$ [样本容量乘以第（ii）步所得到的 R-平方]。

(iv) 将 LM 与 χ_q^2 分布中适当的临界值 c 相比较，如果 $LM > c$，就拒绝虚拟假设。最好能得到 p 值，即 χ_q^2 随机变量超过检验统计量的值的概率。如果小于理想的显著性水平，那么就拒绝 H_0；否则，我们就不能拒绝 H_0。拒绝法则在本质上与 F 检验如出一辙。

根据其形式，LM 统计量有时也被称为 ***n*-*R*-平方统计量**（n-R-squared statistic）。与 F 统计量不同，无约束模型中的自由度在进行 LM 检验时没有什么作用。所有起作用的因素只是被检验约束的个数（q）、辅助回归 R-平方的大小（R_u^2）和样本容量（n）。无约束模型中的 df 没有什么作用是因为 LM 统计量的渐近性质。但我们必须确定将 R_u^2 乘以样本容量以得到 LM；如果 n 很大，R_u^2 看上去较低的值仍可能导致联合显著性。

在举例之前，要警告一点。在第（i）步中，如果我们错误地将 y 对所有自变量进行回归，并将从这个无约束回归中得到的残差用于第（ii）步，那我们就不可能得到一个有意义的统计量：由此得到的 R-平方将恰好为零！这是因为 OLS 选择的估计值使得残差在样本中与所有的自变量都不相关[参见方程（3.13）]。因此，我们只能通过将受约束的残差对所有自变量进行回归来检验式（5.12）。（将受约束残差对受约束的自变量集进行回归也将导致 $R^2=0$。）

173 **例 5.3 犯罪的经济模型**

我们通过对例 3.4 的犯罪模型略加扩展来说明 LM 检验：

$$narr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 tottime + \beta_4 ptime86 \\ + \beta_5 qemp86 + u$$

式中，*narr*86 为一个人被拘捕的次数；*pcnv* 为以前被拘捕后被定罪的次数；*avgsen* 为过去定罪后被判刑的平均时间长度；*tottime* 为此人在年龄达到 18 岁后在 1986 年以前被送进监狱的总次数；*ptime*86 为 1986 年坐牢的月数；*qemp*86 为此人在 1986 年合法就业的季度数。我们要使用 *LM* 统计量检验的虚拟假设是，在控制了其他因素后，*avgsen* 和 *tottime* 对 *narr*86 没有影响。

在第（i）步，我们通过将 *narr*86 对 *pcnv*，*ptime*86 和 *qemp*86 进行回归来估计受约束模型；回归中排除了变量 *avgsen* 和 *tottime*。我们从这个回归中得到 2 725 个残差 $\tilde{u}$。接下来，我们将

$$\tilde{u} \text{ 对 } pcnv, avgsen, tottime, ptime86 \text{ 和 } qemp86 \text{ 进行回归} \tag{5.15}$$

和平常一样，我们列出自变量的顺序是无所谓的。第 2 个回归产生了 R_{u}^{2}，结果其大小约为 0.001 5。这可能看起来很小，但我们必须将它乘以 n 才能得到 *LM* 统计量：$LM=2\,725\ (0.001\,5)\approx 4.09$。自由度为 2 的 χ 平方分布在显著性水平为 10% 时的临界值是 4.61（保留到小数点后两位；参见表 G.4）。于是，我们在 10% 的水平上不能拒绝虚拟假设 $\beta_{avgsen}=0$ 和 $\beta_{tottime}=0$。因为 p 值等于 $P\ (\chi_{2}^{2}>4.09)\approx 0.129$，所以我们在 15% 的水平上将拒绝 H_0。

作为比较，对 *avgsen* 和 *tottime* 联合显著性的 F 检验所得到的 p 值约等于 0.131，它与使用 *LM* 统计量所得到的 p 值相当接近。这无足为奇，因为这两个统计量犯第 1 类错误的概率（渐近地）相同。（即在虚拟假设正确时，它们以相同的频率拒绝虚拟假设。）

像前面这个例子所表明的那样，在大样本的情况下，我们几乎看不到 *LM* 和 F 检验的结果之间有什么重大分歧。由于绝大多数回归软件包都例行计算了 F 统计量，所以我们以后主要使用 F 统计量。但当应用研究中用到 *LM* 统计量时，你应该清楚它。

对 *LM* 统计量最后一个评注是，像对 F 统计量的讨论一样，你必须确定在第（i）和第（ii）步中使用了同样的观测。如果在虚拟假设下被排除的某些自变量存在数据缺失的问题，那么第（i）步得到的残差应该来自于对削减后的数据集的回归。

5.3 OLS 的渐近有效性

我们知道，在高斯-马尔科夫假定下，OLS 估计量是最优线性无偏的。在高斯-马尔科夫假定下，OLS 在一类估计量中也是**渐近有效的**（asymptotically efficient）。要对这个问题做一般性的处理比较困难［参见 Wooldridge (1999，第 4 章)］。现在，我们只在简单回归的情形中描述这个结论。 174

在模型

$$y = \beta_0 + \beta_1 x + u \tag{5.16}$$

中，在假定 MLR.3 下，u 的条件均值为零：$E(u|x)=0$。这样就可以得到一系列 β_0 和 β_1 的一致估计量；如往常一样，我们关注斜率参数 β_1。令 $g(x)$ 为 x 的任意一个函数；比如 $g(x)=x^2$ 或 $g(x)=1/(1+|x|)$。那么 u 就与 $g(x)$ 无关（参见附录 B 中的性质 CE.5）。对所有的观测 i，令 $z_i = g(x_i)$。假定 $g(x)$ 和 x 相关 [记住，因为相关度量了线性依赖，所以有 $g(x)$ 和 x 可能不相关]，那么估计量

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \bar{z})y_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i} \tag{5.17}$$

就是对 β_1 的一致估计。为了看出这一点，我们将 $y_i = \beta_0 + \beta_1 x_i + u_i$ 代入，并把 $\tilde{\beta}_1$ 写成

$$\tilde{\beta}_1 = \frac{\beta_1\left[n^{-1}\sum_{i=1}^{n}(z_i - \bar{z})u_i\right]}{n^{-1}\sum_{i=1}^{n}(z_i - \bar{z})x_i} \tag{5.18}$$

现在，我们可以在分子和分母中应用大数定律，它们分别收敛于 $\mathrm{Cov}(z,u)$ 和 $\mathrm{Cov}(z,x)$。给定 $\mathrm{Cov}(z,u)\neq 0$（所以 z 和 x 相关），由于在假定 MLR.3 下 $\mathrm{Cov}(z,u)=0$，所以我们有

$$\mathrm{plim}\,\tilde{\beta}_1 = \beta_1 + \mathrm{Cov}(z,u)/\mathrm{Cov}(z,x) = \beta_1$$

证明 $\tilde{\beta}_1$ 是渐近正态的，就更困难了。不过，使用与附录中类似的论证，可以证明 $\sqrt{n}(\tilde{\beta}_1 - \beta_1)$ 是渐近正态的，其均值为零，渐近方差为 $\sigma^2\mathrm{Var}(z)/[\mathrm{Cov}(z,x)]^2$。OLS 估计量的渐近方差在 $z=x$ 时得到，在这种情况下，$\mathrm{Cov}(z,x)=\mathrm{Cov}(x,x)=\mathrm{Var}(x)$。因此，$\sqrt{n}(\hat{\beta}_1 + \beta_1)$* 的渐近方差（其中 $\hat{\beta}_1$ 是 OLS 估计量）是 $\sigma^2\mathrm{Var}(x)/[\mathrm{Var}(x)]^2 = \sigma^2/\mathrm{Var}(x)$。现在，柯西-施瓦茨不等式（参见附录 B.4）意味着，$[\mathrm{Cov}(z,x)]^2 \leqslant \mathrm{Var}(z)\mathrm{Var}(x)$，这意味着 $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ 的渐近方差不比 $\sqrt{n}(\tilde{\beta}_1 - \beta_1)$ 的渐近方差大。我们已经在高斯-马尔科夫假定下，针对简单回归情形证明：OLS 估计量的渐近方差比形如式（5.17）的任何一个估计量的渐近方差都小。[式（5.17）中的估计量是工具变量估计量的一个例子，我们在第 15 章将会广泛研究工具变量估计量的问题。] 如果同方差性假定不成立，那么就有一些形如式（5.17）的估计量具有比 OLS 更小的渐近方差。我们在第 8 章将看到这一点。

一般情形与此相似，但在数学上则要困难得多。在 k 个回归元的情形中，将 OLS 的一阶条件推广，可以得到一类一致估计量：

---

* 原书误为 $\sqrt{n}(\tilde{\beta}_1 - \beta_1)$。——译者注

175

$$\sum_{i=1}^{n} g_j(x_i)(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_k x_{ik}) = 0, j = 0,1,\cdots,k \quad (5.19)$$

式中，$g_j(x_i)$ 为第 i 次观测的所有自变量的任意函数。就像从式（5.19）与 OLS 一阶条件式（3.13）的比较中看到的那样，当 $g_0(x_i)=1$ 和对 $j=1,2,\cdots,k$，$g_j(x_i)=x_{ij}$时，我们得到 OLS 估计量。由于我们可以使用我们想使用的 x_{ij}的任意函数，所以式（5.19）中的估计量是无限的。

定理 5.3（OLS 的渐近有效性）

在高斯-马尔科夫假定下，令 $\tilde{\beta}_j$ 表示从求解式（5.19）所得到的估计量，而 $\hat{\beta}_j$ 表示 OLS 估计量。那么，对 $j=0,1,2,\cdots,k$，OLS 估计量具有最小的渐近方差：$\text{Avar}\sqrt{n}(\hat{\beta}_j-\beta_j)\leqslant\text{Avar}\sqrt{n}(\tilde{\beta}_j-\beta_j)$。

更不用说证明它们的渐近正态性了。单纯证明式（5.19）中估计量的一致性，在数学上就是困难的。[参见 Wooldridge（1999，第 5 章）。]

▶ 小　结

作为本章材料背景所申述的条件是相当技术性的，但其实际含义却又很简单。我们已经证明了前四个高斯-马尔科夫假定意味着 OLS 是一致的，而且无须假定误差是从一个正态分布中抽取的（等于是说给定解释变量下 y 的分布不是正态的），我们就知道在第 4 章了解到的所有检验和构造置信区间的方法都是渐近有效的。这意味着，甚至在因变量并不近似于正态分布的情况下，我们也可以应用 OLS，并在一系列应用中使用以前的检验方法。我们还证明了，LM 统计量可以取代 F 统计量用于排除性约束的检验。

在结束本章以前，像例 5.3 那样的例子的确具有一些我们必须特别关注的问题。像 *narr*86 那样的一个变量，它对于总体中的多数人来说都是 0 或 1，一个线性模型可能不足以刻画它与解释变量之间的函数关系。而且，即使一个线性模型能够描述预期的拘捕次数，也可能存在异方差性的问题。这种问题不会随着样本容量的扩大而消失，我们在以后的章节中还会继续讨论这个问题。

关键术语

渐近偏误
渐近置信区间
渐近正态性
渐近性质

辅助回归
一致性
不一致性
拉格朗日乘数 LM 统计量

渐近标准误
渐近 t 统计量
渐近方差
渐近有效性
大样本性质
n-R-平方统计量
得分统计量

习 题

176 5.1 在满足假定 MLR.1－MLR.4 的简单回归中，我们证明了斜率估计量 $\hat{\beta}_1$ 是 β_1 的一致估计。利用$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1$ 证明 $\text{plim}\hat{\beta}_0 = \beta_0$。[你在使用 $\beta_0 = \text{E}(y) - \beta_1(x_1)$的同时，还需要使用 $\hat{\beta}_1$ 的一致性和大数定律。]

5.2 假设模型

$$pctstck = \beta_0 + \beta_1 funds + \beta_2 risktol + u$$

满足前四个高斯-马尔科夫假定，其中 *pctstck* 表示工人养老金投资于股票市场的百分比；*funds* 表示工人可以选择的共同基金的个数；而 *risktol* 表示对风险承受能力的某种度量（*risktol* 越大，则表明这个人对风险的承受能力越强）。如果 *funds* 和 *risktol* 正相关，*pctstck* 对 *funds* 简单回归的斜率系数 $\tilde{\beta}_1$ 有怎样的不一致性？

5.3 数据集 SMOKE.RAW 包含有美国成人个人随机样本在吸烟行为和其他变量方面的信息。变量 *cigs* 是（平均）每天吸烟的数量。你是否认为在美国这个总体中，*cigs* 具有正态分布？试做解释。

5.4 在简单回归模型（5.16）中，我们在前四个高斯-马尔科夫假定下证明了，形如式（5.17）的估计量是斜率 β_1 的一致估计量。给定这样的一个估计量，定义 β_0 的一个估计量为 $\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1\bar{x}$。证明 $\text{plim}\,\tilde{\beta}_0 = \beta_0$。

计算机习题

5.5 本题使用 WAGE1.RAW 中的数据。

(i) 估计方程

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

保存残差并画出其直方图。

(ii) 以 log（*wage*）作为因变量重做第（i）部分。

(iii) 你认为水平值—水平值模型抑或对数—水平值模型更近于满足假定 MLR.6。

5.6 本题使用 GPA2.RAW 中的数据。

(i) 使用所有 4 137 个观测，估计方程

$$colgpa = \beta_0 + \beta_1 hsperc + \beta_2 sat + u$$

并以标准形式报告结论。

(ii) 使用前 2 070 个观测再重新估计第 (i) 部分中的方程。

(iii) 求出第 (i) 部分与第 (ii) 部分所得到的标准误的比率，并将这个比率与式 (5.10) 中的结论相比较。

5.7　在第 4 章的方程 (4.42) 中，计算 *LM* 统计量以检验 *motheduc* 和 *fatheduc* 是否联合地显著。在计算受约束模型的残差时，注意估计受约束模型所使用的观测结果，必须限于在不受约束模型中有全部变量数据可供使用的那些观测 (参见例 4.9)。

附录 5A

177　我们在简单回归情形中，大致勾勒出对 OLS 渐近正态的一个证明 [定理 5.2 (i)]。将简单回归模型写成式 (5.16)。于是，通过对简单回归进行通常的代数运算，我们就能得到

$$\sqrt{n}(\hat{\beta}_1-\beta_1)=(1/s_x^2)\left[n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i\right]$$

式中，s_x^2 为 $\{x_i:\ i=1,\ 2,\ \cdots,\ n\}$ 的样本方差。根据大数定律 (参见附录 C)，$s_x^2 \xrightarrow{P} \sigma_x^2=\mathrm{Var}(x)$。假定 MLR.4 排除了完全共线性，这意味着 $\mathrm{Var}(x)>0$(x_i 在样本中是变化的，因此 x 在总体中就不是一个常数)。接下来，$n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i=n^{-1/2}\sum_{i=1}^{n}(x_i-\mu)u_i+(\mu-\bar{x})\left[n^{-1/2}\sum_{i=1}^{n}u_i\right]$，其中 $\mu=\mathrm{E}(x)$ 是 x 的总体均值。现在，$\{u_i\}$ 就是由均值为零和方差为 σ^2 的 *i.i.d.* 随机变量构成的一个序列，所以 $n^{1/2}\sum_{i=1}^{n}u_i$ 随着 $n\rightarrow\infty$ 而收敛于 $\mathrm{Normal}(0,\sigma^2)$；这正是附录 C 中的中心极限定理。根据大数定律，$\mathrm{plim}(\mu-\bar{x})=0$。渐近理论的一个标准结论是，如果 $\mathrm{plim}(w_n)=0$，而 z_n 具有渐近的正态分布，那么 $\mathrm{plim}(w_nz_n)=0$。[更详细的结论可参见伍德里奇 (Wooldridge，1999，第 3 章)。] 这意味着 $(\mu-\bar{x})(n^{-1/2}\sum_{i=1}^{n}u_i)$ 的概率极限为零。于是，$\{(x_i-\mu)u_i:\ i=1,2,\cdots\}$ 也是一个 *i.i.d.* 随机变量的序列，这些随机变量的均值为零——因为在假定 MLR.3 下 u 和 x 不相关，方差为 $\sigma^2\sigma_x^2$——根据同方差性假定 MLR.5。因此，$n^{-1/2}\sum_{i=1}^{n}(x_i-\mu)u_i$ 具有一个渐近的 $\mathrm{Normal}(0,\sigma^2\sigma_x^2)$ 分布。我们刚刚证明了，$n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i$ 与 $n^{-1/2}\sum_{i=1}^{n}(x_i-\mu)u_i$ 之差的概率极限为零。渐近理论的一个结论是，如果 z_n 具有渐近的正态分布，而且 $\mathrm{plim}(v_n-z_n)=0$，那么 v_n 就具有一个和 z_n 一

样的渐近正态分布。于是，$n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i$ 也具有一个渐近的 Normal（0，$\sigma^2\sigma_x^2$）分布。综合这些结论就得到

$$\sqrt{n}(\hat{\beta}_1-\beta_1)=(1/\sigma_x^2)\left[n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i\right]+\left[(1/s_x^2)\right.$$
$$\left.(1/\sigma_x^2)n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i\right]$$

而由于 plim（$1/s_x^2$）$=1/\sigma_x^2$，所以第二项的概率极限为零。因此，$\sqrt{n}(\hat{\beta}_1-\beta_1)$的渐近分布就是 Normal(0，$\{\sigma^2\sigma_x^2\}/\{\sigma_x^2\}^2$) = Normal($0,\sigma^2/\sigma_x^2$)。由于在简单回归情形中，$\sigma^2=\sigma_x^2$，所以这就完成了在简单回归情形中的证明。至于一般情形，可参见伍德里奇的著述（Wooldridge，1999，第 4 章）。

第 6 章　多元回归分析：其他问题

178　本章将多元回归分析中的几个问题集中到一起，这样的问题我们在前面的章节中都不能很方便地讨论。虽然这些专题不像第 3 章和第 4 章中的内容那么基本，但在将多元回归应用于广泛的实证问题时却具有重要地位。

6.1　数据的测度单位对 OLS 统计量的影响

我们在第 2 章的双变量回归中简要地讨论了改变度量单位对 OLS 截距和斜率估计值的影响，还证明了改变度量单位并不影响 R-平方。我们现在回到数据测度单位的问题，并考察改变因变量或自变量的测度单位对标准误、t 统计量、F 统计量和置信区间的影响。

我们将讨论所有我们预期要发生而又实际发生的情况。当对变量重新测度时，系数、标准误、置信区间、t 统计量和 F 统计量改变的方式，都不影响所有被测度的影响和检验结果。尽管这无足为奇（实际上如果不是这样，我们才真正担心），但看一下到底有什么变化，大有好处。怎样度量数据通常只起非实质性的作用，比如说，减少所估计系数中小数点后零的个数等。通过对度量单位英明的选择，我们可以在不做任何本质改变的情况下，改进

所估计方程的表现。

虽然我们也可以用一般的方式来看待这个问题，但使用例子更便于说明。同样，这里引入抽象的记号并没有什么价值。

我们从将婴儿出生体重与孕妇吸烟量和家庭收入的一个关系式开始：

$$\widehat{bwght} = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 faminc \tag{6.1}$$

式中，*bwght* 为以盎司为单位的孩子出生时体重；*cigs* 为母亲在怀孕期间每天吸烟的数量；*faminc* 为以千美元为单位的家庭年收入。使用BWGHT.RAW 中数据所得到这个方程的估计值，见于表 6.1 中的第（1）列。标准误放在括号中。*cigs* 的估计值说明，如果一个妇女每天多吸 5 支烟，那么预计婴儿出生体重约减少0.463 4(5) = 2.317 盎司。因为 *cigs* 的 *t* 统计量是 −5.03，所以这个变量是极为统计显著的。

表 6.1　　数据测度的影响

179

因变量	(1) *bwght*	(2) *bwghtlbs*	(3) *bwght*
自变量			
cigs	−0.463 4 (0.091 6)	−0.028 9 (0.005 7)	—
packs	—	—	−9.268 (1.832)
faminc	0.092 7 (0.029 2)	0.005 8 (0.001 8)	0.092 7 (0.029 2)
intercept	116.974 (1.049)	7.310 9 (0.065 6)	116.974 (1.049)
观测个数	1 388	1 388	1 388
R-平方	0.029 8	0.029 8	0.029 8
SSR	557 485.51	2 177.677 8	55 748 5.51
SER	20.063	1.253 9	20.063

现在，假使我们决定以磅而不是以盎司为单位来度量出生体重。令 *bwghtlbs* = *bwght* /16 为出生体重的磅数。如果我们在估计方程中以这个变量为因变量，我们的 OLS 统计量会怎么样呢？通过对方程（6.1）的简单计算，很容易就能发现其对系数估计值的影响。将整个方程（6.1）的两边同时除以 16，得到

$$\widehat{bwght}/16 = \hat{\beta}_0/16 + (\hat{\beta}_1/16)cigs + (\hat{\beta}_2/16)faminc$$

由于方程的左边就是出生体重的磅数，于是每一个新的系数都对应于老系数的 1/16。为了验证这一点，将 *bwghtlbs* 对 *cigs* 和 *faminc* 的回归报告在表 6.1 的第（2）列。在保留四位有效数字的情况下，第（2）列中的截距和

斜率刚好就是第(1)列中对应数据除以16。比如，*cigs* 的系数现在是 $-0.028\,9$；这意味着，如果 *cigs* 提高5支，出生体重将减少 $0.028\,9(5) =$
180 $0.144\,5$ 磅。以盎司为单位，则减少 $0.144\,5(16) = 2.312$ 盎司，与我们前面得到的2.32盎司只是因为保留小数而略有差异。要点是，无论如何度量因变量，一旦把影响转化成同样的度量单位，我们就会得到同样的结论。

统计显著性怎么样呢？恰如所料，将因变量的单位从盎司改变为磅，对于自变量在统计上的重要性没有任何影响。第（2）列中的标准误刚好就是第（1）列中标准误的1/16。经过一些快速的计算即表明，第（2）列中的 t 统计量实际上等于第（1）列*中的 t 统计量。第（2）列中置信区间的端点恰好也是第（1）列中对应端点除以16。这是因为，CI与标准误变化同样的倍数。[记住，这里95%的CI就是 $\hat{\beta}_j \pm 1.96\mathrm{se}(\hat{\beta}_j)$。]

用拟合优度的术语来说，从两个回归中所得到的 R-平方理所当然也是一样的。注意，两个方程的残差平方和 *SSR* 和回归标准误 SER 确实有所不同。这些差异很容易解释。令 $\hat{u}_i$ 表示原方程（6.1）中第 i 个观测的残差，于是当因变量为 *bwghtlbs* 时，残差仅仅是 $\hat{u}_i/16$。因此，第二个方程的残差平方是 $(\hat{u}_i/16)^2 = \hat{u}_i^2/256$。这就是为什么第（2）列的残差平方和等于第（1）列中的SER除以256。

由于 $\mathrm{SER} = \hat{\sigma} = \sqrt{\mathrm{SSR}/(n-k-1)} = \sqrt{\mathrm{SSR}/1\,385}$，所以第（2）列中的SER就是第（1）列中SER的1/16。对此的另一种思考方法是，在以 *bwghtlbs* 为因变量的方程中，其误差的标准差只是原方程误差的标准差的1/16。这并不意味着我们已经通过改变出生体重的度量方法而减小了误差；较小的SER仅仅反映了度量单位的差异。

接下来，让我们把因变量换回它原来的度量单位：以盎司为单位的 *bwght*。相反，让我们改变自变量之一 *cigs* 的度量单位。定义 *packs* 为每天吸烟的包数，于是 $packs = cigs/20$。现在系数和其他OLS统计量会有什么变化呢？于是，我们可以写出

$$\widehat{bwght} = \hat{\beta}_0 + (20\hat{\beta}_1)(cigs/20) + \hat{\beta}_2 faminc = \hat{\beta}_0 + (20\hat{\beta}_1)packs + \hat{\beta}_2 faminc$$

于是，截距和 *faminc* 的系数都没有变化，只是 *packs* 的系数是 *cigs* 的系数的20倍。这在直觉上很具吸引力。*bwght* 对 *packs* 和 *faminc* 进行回归的结论，列在表6.1的第（3）列中。顺便提一句，记住在同一个方程中同时包括 *cigs* 和 *packs* 是毫无意义的：这样将导致完全共线性，并没有任何实质性的意义。

问题6.1

在原出生体重方程（6.1）中，假设是以美元而不是以千美元来度量。因而，定义变量=1 000，当用 *faminc* 代替 *fincdol* 时，OLS统计量将如何变化？为便于报告回归结论起见，你认为是用美元度量收入更好，还是用千美元度量收入更好？

---

*原书误为第（2）列。——译者注

除了 *packs* 的系数外，第（3）列中还有另一个统计量与第（1）列中不同：*packs* 的标准误是第（1）列中 *cigs* 的标准误的 20 倍。这意味着无论我们是用支还是用包来度量吸烟量，检验吸烟显著性的 t 统计量都是一样的。这只是自然而然的。

上例详细地阐述了重新测度因变量和自变量所引起的多种可能性。对经济学中的美元数字，特别是在美元数字很大时，通常需要进行重新测度。

181 我们在第 2 章证明了，如果因变量以对数形式出现，那么改变度量单位就不会影响斜率系数。这里同样如此：当因变量以对数形式出现时，改变因变量的度量单位仍不会影响任何一个斜率估计值。这是基于如下简单的事实：对于任何一个常数 $c_1>0$，都有$\log(c_1y_i)=\log(c_1)+\log(y_i)$。新的截距将是$\log(c_1)+\hat{\beta}_0$。与此类似，对任何一个 x_j，当它在回归中以$\log(x_j)$出现时，改变其度量单位只能影响到截距。这与我们对百分比变化和（特别是）弹性的了解相对应：它们不会随着 y 或 x_j 度量单位的变化而变化。比如，如果我们设定式(6.1)中的因变量为 log(*bwght*)，估计这个方程，然后在因变量为log(*bwghtlbs*)的情况下重新估计它，那么，在这两个回归中，*cigs* 和 *faminc* 的系数将相同，只是截距略有不同。

β 系数

在计量经济的应用中，有时会采用一个难以解释的尺度来度量一个关键变量。劳动经济学家常常在工资方程中包括考试分数，而对这些考试给出分数时常常很随意而又难以解释。（至少对经济学家来说如此！）在几乎所有情形中，我们感兴趣的，只是将个人得分与总体相比较。因此，不是问（比方说）考试得分提高 10 分对小时工资有什么影响，问考试分数提高 1 倍的标准差会怎么样将更有意义。

假定我们已经得到样本标准差（在多数回归软件包中都很容易），当所估计模型中的一个自变量提高其标准差的一定倍数时，我们轻而易举就能看到其对因变量的影响。这通常是一个好主意。所以，如果在我们看标准化考试分数（比如 *SAT* 分数）对大学 *GPA* 的影响时，我们就能求出 *SAT* 的标准差，并看 *SAT* 分数提高 1 倍或 2 倍标准差又会怎么样。

有时候，在包括因变量和全部自变量在内的所有变量都被标准化的情况下，得到回归的结论也有用处。在样本中进行标准化的过程，就是将一个变量减去其均值，然后除以其标准差（参见附录 C）。这意味着，我们对样本中的每一个变量都计算 z-得分。然后，我们用这些 z-得分进行回归。

为什么标准化很有用呢？最容易的办法是从变量都保持其原有形式的 OLS 方程开始：

$$y_i=\hat{\beta}_0+\hat{\beta}_1x_{i1}+\hat{\beta}_2x_{i2}+\cdots+\hat{\beta}_kx_{ik}+\hat{u}_i \tag{6.2}$$

我们已经在方程中包括了表示观测次数的下标 i，以强调标准化适用于所有

的样本值。现在，如果对式（6.2）求出其平均方程，利用 $\hat{u}_i$ 具有零样本均值的事实，并将式（6.2）减去平均方程，就得到

$$y_i - \bar{y} = \hat{\beta}_1(x_{i1} - \bar{x}_1) + \hat{\beta}_2(x_{i2} - \bar{x}_2) + \cdots + \hat{\beta}_k(x_{ik} - \bar{x}_k) + \hat{u}_i$$

182 现在，令 $\hat{\sigma}_y$ 为因变量的样本标准差，$\hat{\sigma}_1$ 为 x_1 的样本标准差，$\hat{\sigma}_2$ 为 x_2 的样本标准差，等等。然后简单的运算就得到方程

$$\begin{aligned}(y_i - \bar{y})/\hat{\sigma}_y = (\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1[(x_{i1} - \bar{x}_1)/\hat{\sigma}_1] + \cdots \\ + (\hat{\sigma}_k/\hat{\sigma}_y)\hat{\beta}_k[(x_{ik} - \bar{x}_k)\hat{\sigma}_k] + (\hat{u}_i/\hat{\sigma}_y)\end{aligned} \tag{6.3}$$

式（6.3）中的每个变量都用其 z-得分而被标准化，这就得到一些新的斜率参数。比如，$(x_{i1} - \bar{x}_1)\hat{\sigma}_1$ 的斜率参数为 $(\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1$。这个系数也就是原系数 $\hat{\beta}_1$ 乘上 x_1 的标准差与 y 的标准差之比。截距项则完全消失。

改写式（6.3）很有帮助，省略下标 i 后，为

$$z_y = \hat{b}_1 z_1 + \hat{b}_2 z_2 + \cdots + \hat{b}_k z_k + \text{误差} \tag{6.4}$$

式中，z_y 为 y 的 z-得分；z_1 为 x_1 的 z-得分，等等。新的系数是

$$\hat{b}_j = (\hat{\sigma}_j/\hat{\sigma}_y)\hat{\beta}_j, \quad j = 1, \cdots, k \tag{6.5}$$

传统上称这些 $\hat{b}_j$ 为**标准化系数**（standardized coefficients）或 **β 系数**（beta coefficients）。（后面这个名称更常见，但不巧的是，我们已经用 $\hat{\beta}$ 表示常用的 OLS 估计值了。）

方程（6.4）赋予 β 系数颇有意思的含义：如果 x_1 提高 1 倍的标准差，那么 $\hat{y}$ 就变化 $\hat{b}_1$ 倍的标准差。于是，我们不是在以 y 或 x_j 的原有单位来度量其影响，而是以标准差为单位。由于它使得回归元的测度无关紧要，所以这个方程把所有的解释变量放到相同的地位上。在一个标准的 OLS 方程中，不可能只看不同系数的大小，也不可能断定具有最大系数的解释变量就“最重要”。我们在前面看到，通过改变 x_j 的度量单位，可以任意改变系数的大小。但当每个 x_j 都被标准化之后，比较由此得到的 β 系数就更加有说服力。

我们总可以先将 y，x_1，…，x_k 标准化，然后将 y 的 z-得分对 x_1，…，x_k 的 z-得分进行 OLS 回归，从而得到 β 系数，其中由于截距项将为零，所以不一定要包括一个截距项。有些回归软件包通过一个简单的命令就能给出 β 系数。如下例子就说明了 β 系数的用处。

例 6.1　污染对住房价格的影响

我们使用例 4.5 中的数据（在文件 HPRICE2.RAW 中）来说明 β 系数的用处。记得 *nox* 是关键的自变量，它是对每个社区空气中的氧化亚氮含

183 量的一种度量。理解污染效应大小的方法之一（无须过问氧化亚氮对空气质量影响的基础科学知识），就是去计算 β 系数。（另一种方法包含在例 4.5 中：我们通过使用 *price* 和 *nox* 的对数形式而得到价格对 *nox* 的弹性。）

总体方程是一个水平值—水平值的模型：

$$price = \beta_0 + \beta_1 nox + \beta_2 crime + \beta_3 rooms + \beta_4 dist + \beta_5 stratio + u$$

式中，除了 *crime* 之外的所有变量都和例 4.5 中的定义一样；*crime* 为所报告的人均犯罪次数。β 系数的报导见于如下方程中报告（每个变量都被转换成了其 z-得分）：

$$\widehat{zprice} = -0.340\,znox - 0.143\,zcrime + 0.514\,zrooms$$
$$-0.235\,zdist - 0.270\,zstratio$$

这个方程表明，*nox* 提高 1 倍的标准差，会使价格下降 0.34 倍的标准差；*crime* 提高 1 倍的标准差，会使价格减少 0.14 倍的标准差。因此，总体中同样的相对变化，污染比犯罪对住房价格产生更大的影响。用房间数（*rooms*）度量的住房大小，则具有最大的标准化影响。如果我们想知道每个自变量对平均住房价格之美元价值影响，那我们就应该使用未经标准化的变量。

6.2 对函数形式的进一步讨论

在前面几个例子中，我们已经遇到了在计量经济学中容许被解释变量和解释变量之间出现非线性关系最普遍的技巧：使用因变量或自变量的对数形式。我们还看到了一些包含某些自变量平方项的模型，但我们仍未对它们进行系统处理。在本节中，我们讨论对函数形式的一些变形和推广，而这种变形和推广在应用研究中又经常出现。

对使用对数函数形式的进一步讨论

首先，我们回忆一下如何解释模型

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 rooms + u \tag{6.6}$$

中的参数，其中的变量都取自例 4.5。记住，全书的 $\log(x)$ 都表示 x 的自然对数。系数 β_1 就是 *price* 对 *nox*（污染）的弹性。系数 β_2 是在 $\Delta rooms = 1$ 时 $\log(price)$ 的变化；就像我们多次看到的那样，把它乘以 100 就近似等于价格变化的百分数。记得 $100 \cdot \beta_2$ 有时也被称为 *price* 对 *rooms* 的半弹性。

当使用 HPRICE2.RAW 中的数据进行估计时，我们得到

$$\begin{aligned} \log(\widehat{price}) &= \underset{(0.19)}{9.23} - \underset{(0.066)}{0.718}\log(nox) + \underset{(0.019)}{0.306}\,rooms \\ n &= 506,\ R^2 = 0.514 \end{aligned} \tag{6.7}$$

184 因此，当 *nox* 提高 1% 时，*price* 在保持 *rooms* 不变的情况下会下降 0.718%。当 *rooms* 增加 1 时，会提高近 100（0.306）= 30.6%。

这个应用研究最终表明，多一个房间会提高价格近 30.6% 的估计值，

多少有些不太准确。因为随着 $\log(y)$ 的变化变得越来越大，$\%\Delta y \approx 100 \cdot \Delta\log(y)$ 的近似就会越来越不准确，所以就出现了近似误差。幸运的是，可以使用一种简单的运算就能计算出精确的百分比变化。

为了描述这种程序，我们考虑如下一般估计模型：

$$\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2$$

（增加其他自变量不致改变这个程序。）现在，固定 x_1，我们有 $\Delta\widehat{\log(y)} = \hat{\beta}_2 \Delta x_2$。使用指数函数和对数函数的简单数学性质，可给出所预计的 y 的精确百分比变化为

$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_2 \Delta x_2) - 1] \tag{6.8}$$

乘以 100 后，就将比例变化转化成了百分比变化。当 $\Delta x_2 = 1$ 时，

$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_2) - 1] \tag{6.9}$$

以 $x_2 = rooms$ 和 $\hat{\beta}_2 = 0.306$ 应用于住房价格一例，得到 $\%\hat{\Delta}(\widehat{price}) = 100[\exp(0.306) - 1] = 35.8\%$，它明显大于从式（6.7）中直接得到的近似百分比变化 30.6%。[顺便一提，由于 $\exp(\cdot)$ 是一个非线性函数，所以它不是一个无偏估计，但它是 $100[\exp(\beta_2) - 1]$ 的一个一致估计量。这是因为概率极限可以通过连续函数变换，而期望值算子却不能。参见附录 C。]

方程（6.8）中的调整对于小的百分比变化而言并没有那么重要。比如，当我们在方程（6.7）中包括学生—教师比，那么其估计系数便是 −0.052，意味着如果将 *stratio* 提高 1，则近似下降 5.2%。精确的比例变化为 $\exp(-0.052) - 1 \approx -0.051$ 或 −5.1%。另一方面，如果我们将 *stratio* 提高 5，那么价格的近似百分比变化就是 −26%，而从方程（6.8）所得到的精确变化为 $100[\exp(-0.26) - 1] \approx -22.9\%$。

我们已经看到，使用自然对数使得对系数的解释颇具吸引力，而且由于斜率系数不随测度单位而变化，所以可以忽略以对数形式出现的变量的度量单位。在应用研究中如此广泛地使用对数，还有其他几方面的原因。首先，当 $y > 0$ 时，使用 $\log(y)$ 作为因变量的模型，通常比使用 y 的水平值作为因变量的模型更接近 CLM 假定。严格为正的变量，其条件分布常常具有异方差性或偏态性；取对数后，即使不能消除这两方面的问题，也可以使之有所缓和。

此外，取对数通常会缩小变量的取值范围，在某些情况下还相当可观。这就使得估计值对因变量或自变量的异常（或极端）观测不是那么敏感。我们在第 9 章专门探讨这种异常观测的问题。

185 至于何时取对数，尽管没有一个固定模式，但也有一些标准的经验法则。对于一个以正的美元数量为单位的变量，通常都可以取对数。对此，我们已经看到诸如工资、薪水、企业销售额和企业的市场价值等变量。像人口、雇员总数和学校注册人数等变量，也常常以对数形式出现；它们具有大正整数的共同特征。

问题 6.2 ☞

假设每年酒后驾驶被拘捕的次数由

$$\log(arrests) = \beta_0 + \beta_1 \log(pop) + \beta_2 age16\sim25 + \text{其他因素}$$

决定，式中 $age16\sim25$ 是年龄在 16 岁～25 岁之间的人口比例。证明 β_2 具有如下（其他条件不变情况下的）解释：它是当年龄在 16 岁～25 岁之间人口的百分比提高 1 个百分点，导致 *arrests* 的百分比变化。

以年度量的变量（如受教育年数、工作经历、任职年限、年龄等），则通常以其原有形式出现。至于比例或百分比变量（如失业率、养老保险金的参与率、学生通过标准化考试的百分比、犯罪报告中的拘捕率等），尽管存在使用其水平值的趋势，但我们既可以使用其原有形式，也可以使用其对数形式。这是因为，任何一个涉及原变量（无论它是因变量还是自变量）的回归系数，都具有一种百分点变化的解释。（对百分比变化和百分点变化之区别的回顾，可参见附录 A。）比方说，如果在回归中使用 log(*unem*)，其中 *unem* 是失业个人占总人口的百分比，我们必须仔细区别百分点变化和百分比变化。记住，如果 *unem* 从 8 变化到 9，就是提高 1 个百分点，但却从原来的失业水平上提高了 12.5%。使用对数意味着，我们在考虑失业率的百分比变化：$\log(9) - \log(8) \approx 0.118$ 或 11.8%，也就是对实际提高 12.5% 的对数近似。

使用对数所受到的一个限制是，变量不能取零或负值。但在 y 非负而又可等于零的情形中，有时采用 $\log(1+y)$。除了从 $y=0$ 开始的变化（此时的百分比变化没有定义）外，通常的百分比变化几乎完全保留了百分比变化的解释。一般而言，当数据并非多数为零时，使用 $\log(1+y)$，并把估计值做变量为 $\log(y)$ 时的解释，通常是可以接受的。这样的一个例子是，y 是制造企业总体中平均对每个雇员培训的小时数，只要大多数企业至少对一个工人进行培训。

使用对数形式的因变量的一个缺陷是，更难以预测原变量的值。原模型使我们能预测 $\log(y)$，而不是 y。不过，把对 $\log(y)$ 的预测转变成对 y 的预测是相当容易的（参见 6.4 节）。一个相关的问题是，将 y 作为因变量的模型的 R-平方，与 $\log(y)$ 作为因变量的模型的 R-平方进行比较，是不合逻辑的。它们解释的是不同变量的变异。我们将在 6.4 节讨论，如何计算可比较的拟合优度的度量指标。

含二次式的模型

在应用经济学中，为了描述递减或递增的边际效应，也常常用到**二次函数**（quadratic functions）。你或许愿意在附录 A 中温习一下二次函数的性质。

186 在最简单的情形中，只取决于单一个观测因素 x，但又取决于其二次形式：

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

比如，取 $y = wage$ 和 $x = exper$。如我们在第 3 章所讨论的那样，这个模型算不上简单回归分析，但又很容易用多元回归来处理。

重要的是要记得，β_1 并没有度量 y 相对 x 的变化；保持 x^2 不变而改变 x 是毫无意义的。如果我们将被估计的方程写成

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \tag{6.10}$$

那么，我们就有如下近似

$$\Delta\hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x)\Delta x \text{，所以 } \Delta\hat{y}/\Delta x \approx \hat{\beta}_1 + 2\hat{\beta}_2 x \tag{6.11}$$

这说明，x 和 y 之间的斜率取决于 x 的值；所估计的斜率是 $\hat{\beta}_1 + 2\hat{\beta}_2 x$。如果代入 $x = 0$，就会看到，$\hat{\beta}_1$ 可被解释为从 $x = 0$ 到 $x = 1$ 的近似斜率。此外，都必须考虑第二项 $2\hat{\beta}_2 x$。

如果我们只对给定 x 的起始值及其变化量时计算 y 的预计变化量感兴趣，那我们就可以直接使用式（6.10）：完全没有理由去使用微积分做近似计算。但我们通常对快捷概括 x 对 y 的影响更感兴趣，而方程（6.11）中对 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的解释就为我们提供了这种概括。通常，我们可以代入样本中 x 的平均值，或某些有意义的数值，比如中位数或上下四分位点。

在多数应用中，$\hat{\beta}_1$ 为正，而 $\hat{\beta}_2$ 则为负。比如，使用 WAGE1.RAW 中的工资数据，我们得到

$$\begin{aligned} wage &= 3.73 + 0.298exper - 0.006\,1exper^2 \\ &\quad\;(0.35)\;(0.041)\qquad\quad (0.000\,9) \\ &\quad\; n = 526,\ R^2 = 0.093 \end{aligned} \tag{6.12}$$

所估计的这个方程意味着 $exper$ 对 $wage$ 具有递减的影响。工作经历的第一年约值每小时 30 美分（0.298 美元）。工作经历的第二年就没那么有价值了[根据式（6.11）在 $x = 1$ 处的近似，约为 $0.298 - 2(0.0061)(1) \approx 0.286$ 或 28.6 美分]。当工作经历从 10 年变化到 11 年时，预计工资只会提高约 $0.298 - 2(0.006\,1)(10) \approx 0.176$ 或 17.6 美分，等等。

当 x 的系数为正，而 x^2 的系数为负时，二项式便具有抛物线形态。总存在一个正的 x 值，此时 x 对 y 的影响为零；在此点之前，x 对 y 的影响为正；而在此点之后，x 对 y 的影响为负。实践中，重要的是要知道这个转折点出现在哪里。

187 在所估计的 $\hat{\beta}_1 > 0$ 和 $\hat{\beta}_2 < 0$ 的方程（6.10）中，转折点（或函数最大值点），总是 x 的系数与 2 倍的 x^2 的系数之比：

$$x^* = |\hat{\beta}_1/(2\hat{\beta}_2)| \tag{6.13}$$

在工资一例中，$x^* = exper^*$ 就是 $0.298/[2(0.006\,1)] \approx 24.4$。（注意，我们在做这个计算时，只须去掉 0.006 1 的负号。）图 6.1 说明了这个二次关系。

在工资方程（6.12）中，工作经历的回报在约24.4年时达到零。我们该怎样解释它呢？至少有三个可能的解释。首先，样本中可能根本就没几个人拥有24年的工作经历，所以曲线在24年以右的部分可以被忽略。使用二次函数来刻画递减效应的代价是，二次函数最终必须转向。如果这个转折点超过了样本中绝大多数人的工作经历，那么，就不值得过多考虑。但在数据集WAGE1.RAW中，样本中约有28%的人具有24年以上的工作经历；这是一个很高的比例，所以不容忽视。

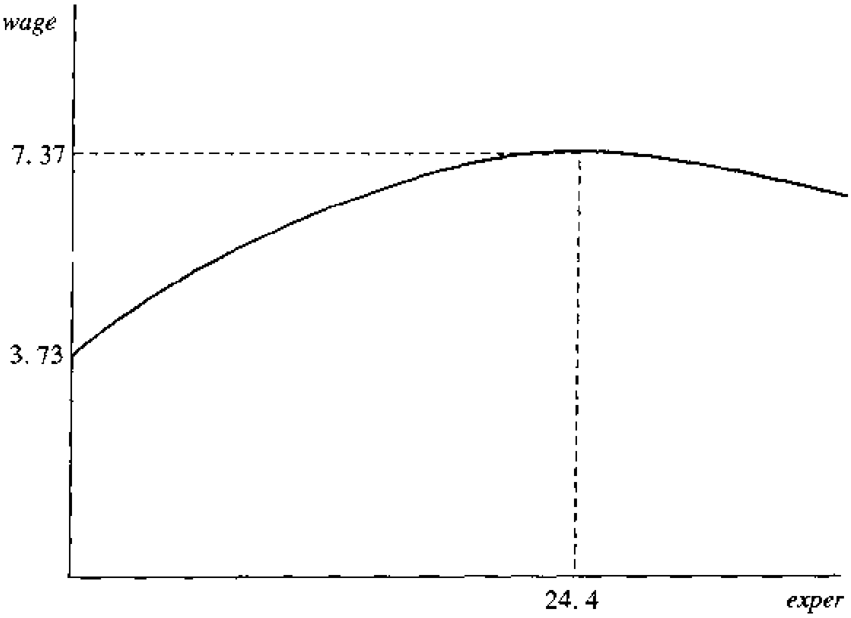


图6.1　*wage*和*exper*之间的二次关系

*exper*的回报确实有可能在到达某点之后变为负，但很难相信它会发生在24年的工作经历上。我们更愿意相信，所估计的*exper*对*wage*的影响是有偏误的，或因为我们没有对其他任何因素进行解释，或因为方程（6.12）中*wage*和*exper*之间的函数关系并不是完全正确。习题6.9要求你除了通过使用log(*wage*)作为因变量外，还要控制受教育程度，来解释上述可能性。 188

当一个模型的因变量是对数形式而解释变量以二次式形式出现时，为了作出一个有用的解释，需要特别小心。下例还表明，二项式也可以具有U形，而不是抛物线形。当方程（6.10）中的$\hat{\beta}_1$为负而$\hat{\beta}_2$为正时，就出现了U形关系，这就刻画了x对y的递增影响。

例6.2　污染对住房价格的影响

我们通过在例4.5中包括进*rooms*的二次项而修改住房价格模型：

$$\log(price) = \beta_0 + \beta_1\log(nox) + \beta_2\log(dist) + \beta_3 rooms + \beta_4 rooms^2 + \beta_5 stratio + u \qquad (6.14)$$

利用HPRICE2.RAW中的数据估计上述模型，得到

$$\log(\widehat{price}) = 13.39 - 0.902\log(nox) - 0.087\log(dist)$$

$$\begin{aligned}&(0.57)\ (0.115)\qquad\quad (0.043)\\&-0.545rooms+0.062rooms^2-0.048stratio\\&(0.165)\qquad\quad (0.013)\qquad\quad (0.006)\\&n=506,\ R^2=0.603\end{aligned}$$

二次项 $rooms^2$ 的 t 统计量约为 4.77，所以它在统计上是相当显著的。但解释 $rooms$ 对 $\log(price)$ 的影响会怎么样呢？起初，其影响看起来有些奇怪。由于 $rooms$ 的系数为负，而 $rooms^2$ 的系数为正，所以这个方程确实意味着在 $rooms$ 的值很低时，多一个房间对 $\log(price)$ 具有负影响。到某个点后，开始变为正影响，这个二次项的形态意味着，$price$ 对 $rooms$ 的半弹性随着 $rooms$ 的增加而递增。这种情况示于图 6.2。

189

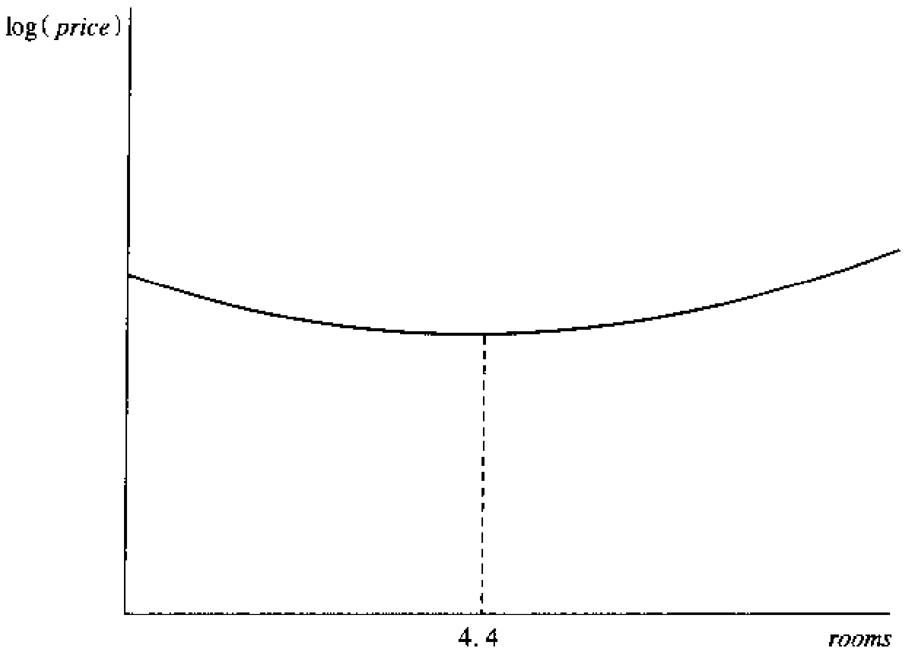


图 6.2　log（*price*）作为 *rooms* 的一个二次函数

我们利用式（6.13）得到 $rooms$ 的转折值（尽管 $\hat{\beta}_1$ 为负而 $\hat{\beta}_2$ 为正）。$rooms$ 系数的绝对值 0.545 除以 $rooms^2$ 系数 0.062 的 2 倍，就得到 $rooms^*=0.545/[2(0.062)]\approx4.4$；这个点在图 6.2 中被标了出来。

我们真的会相信，以 3 间卧室为起点，增加到 4 间卧室实际上会降低一套住房的预期价值吗？可能不会。结果，506 个社区样本中只有 5 个社区住房的平均卧室数不足 4.4 间，约占样本的 1%。二次项在 4.4 左边的部分如此之小，以至实际上可以被忽略。在 4.4 的右边，我们看到，增加一间卧室对价格的百分比变化具有递增的影响：

$$\Delta\widehat{\log}_1(price)\approx\{[-0.545+2(0.062)]rooms\}\Delta rooms$$

所以

$$\begin{aligned}\%\Delta price&\approx100\{[-0.545+2(0.062)]rooms\}\Delta rooms\\&=(-54.5+12.4rooms)\Delta rooms\end{aligned}$$

于是，比方说 *rooms* 从 5 增加到 6 会导致价格提高约 $-54.5+12.4(5)=7.5\%$；*rooms* 从6增加到 7 会导致价格提高约 $-54.5+12.4(6)=19.9\%$。这是一个很强的递增影响。

与对数一起使用二次式还有许多其他的可能情况。比如，允许式（6.14）中的 *price* 和 *nox* 之间有非常数的弹性而将式（6.14）加以扩展，得到

$$\log(price)=\beta_0+\beta_1\log(nox)+\beta_2[\log(nox)]^2+\beta_3 crime$$
$$+\beta_4 rooms+\beta_5 rooms^2+\beta_6 stratio+u \qquad (6.15)$$

190 如果 $\beta_2=0$，那么 β_1 就是 *price* 对 *nox* 的弹性。否则，这个弹性就取决于 *nox* 的水平。为了看出这一点，有

$$\%\Delta price\approx[\beta_1+2\beta_2\log(nox)]\%\Delta nox \qquad (6.16)$$

于是 *price* 对 *nox* 的弹性就是$\beta_1+2\beta_2\log(nox)$，并取决于$\log(nox)$的水平。

最后，回归模型中还可以包括其他的多项式。二次式当然是最常见的，但迟早会出现一个三次或四次式。总成本函数的一个合理的函数形式通常是

$$cost=\beta_0+\beta_1 quantity+\beta_2 quantity^2+\beta_3 quantity^3+u$$

对这样一个模型的估计并没有什么新的困难，只是对参数的解释更难以处理（尽管直接使用了微积分）；所以我们就不再深入研究这些模型。

含有交互作用项的模型

因变量对一个解释变量的偏效应、弹性或半弹性，有时很自然地取决于另一个解释变量的大小。比如，在模型

$$price=\beta_0+\beta_1 sqrft+\beta_2 bdrms+\beta_3 sqrft\cdot bdrms+\beta_4 bthrms+u$$

中，*bdrms* 对 *price* 的偏效应（保持所有其他变量不变）为

$$\frac{\Delta price}{\Delta bdrms}=\beta_2+\beta_3 sqrft \qquad (6.17)$$

若 $\beta_3>0$，则式（6.17）意味着，住房面积越大，增加一间卧室导致价格上升得越多。换言之，住房的平方英尺数与卧室的间数之间存在着**交互效应**（interaction effect）。在概括 *bdrms* 对 *price* 的影响时，我们必须在某些有意义的 *sqrft* 数值（比如样本的均值或上下四分位值）下计算式（6.17）。至于 β_3 是否为零，我们不难检验。

例 6.3　出勤率对期末考试成绩的影响

用听课百分比、先前的大学 *GPA* 和 *ACT* 分数来解释期末考试的标准化成绩的一个模型是

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2$$
$$+ \beta_5 ACT^2 + \beta_6 priGPA \cdot atndrte + u \qquad (6.18)$$

（我们在 6.1 节讨论了使用标准化成绩的原因：更易于解释一个学生相对班上其他同学的成绩。）除了 *priGPA* 和 *ACT* 的平方项之外，该模型还包括了一个 *priGPA* 和出勤率 *atndrte* 的乘积项。其思想是，出勤率对那些过去成 191 绩不同（用 *priGPA* 来度量）的同学具有不同的影响。我们所关心的是出勤率对期末考试分数的影响：$\Delta stndfnl/\Delta atndrte = \beta_1 + \beta_6 priGPA$ 。

利用 ATTEND.RAW 中对学习《微观经济学原理》的学生所观测的 680 个数据，估计出来的方程是

$$\begin{aligned}\widehat{stndfnl} = {} & 2.05 - 0.0067 atndrte - 1.63 priGPA - 0.128 ACT \\ & (1.36)(0.0102) \qquad (0.48) \qquad (0.098) \\ & + 0.296 priGPA^2 + 0.0045 ACT^2 + 0.0056 priGPA \cdot atndrte \\ & (0.101) \qquad (0.0022) \qquad (0.0043) \qquad (6.19) \\ & n = 680,\ R^2 = 0.229,\ \bar{R}^2 = 0.222\end{aligned}$$

我们必须极其小心地解释这个方程。如果我们仅看 *atndrte* 的系数，那就会错误地得出结论：听课对期末考试分数具有负面影响。但这个系数只是度量了 $priGPA = 0$ 时的影响，而 $priGPA = 0$ 的情况又是没有意义的（在这个样本中，最小的 *priGPA* 也约为 0.86）。我们还必须小心，不能仅看 β_1 和 β_6 的估计值就断定：由于每个 t 统计量都不显著，所以不能拒绝 H_0：$\beta_1 = 0$，$\beta_6 = 0$。实际上，这个联合假设的 F 检验的 p 值为 0.014，所以我们在 5% 的显著性水平上，就能肯定地拒绝 H_0。这个例子很好地说明了，在检验一个联合假设时，孤立地看待每个 t 统计量，可能会误入歧途。

该如何估计 *atndrte* 对 *stndfnl* 的偏效应呢？惟有代入有意义的 *priGPA* 值，才能得到其偏效应。样本中 *priGPA* 的均值是 2.59，所以在 *priGPA* 的平均值上，*atndrte* 对 *stndfnl* 的影响是 $-0.0067 + 0.0056(2.59) \approx 0.0078$。其含义是什么呢？由于 *atndrte* 是以百分比度量的，所以它意味着 *atndrte* 提高 10 个百分点，使 *stndfnl* 比期末考试平均分数高出 0.078 倍的标准差。

怎么知道估计值 0.0078 在统计上是否异于零呢？这又要回到回归中来。其中，用 $(priGPA - 2.59) \cdot atndrte$ 来取代 $priGPA \cdot atndrte$，*atndrte* 的新系数就给出了在 $priGPA = 2.59$ 时的估计效应，也给出了其标准误；回归中其他地方都没有变化。（在 4.4 节描述过这种手法。）进行这个新回归，则给出 $\hat{\beta}_1 + \hat{\beta}_6(2.59) = 0.0078$ 的标准误为 0.0026，从而得到 $t = 0.0078/0.0026 = 3$。因此，在 *priGPA* 的均值水平上，我们断定，出勤率对期末考试分数具有统计上显著的影响。

问题 6.3

如果我们在方程（6.18）中增加 $\beta_7 ACT \cdot atndrte$ 一项，那么 *atndrte* 对 *stndfnl* 的偏效应是什么？

因为平方项 $priGPA^2$ 的出现，所以欲得到 $priGPA$ 对 $stndfnl$ 的影响则要复杂得多。为了发现在 $priGPA$ 的均值水平和平均出勤率 0.82 下的影响，我们就用 $(priGPA-2.59)^2$ 取代，并用 $priGPA\cdot(atndrte-0.82)$ 取代 $priGPA\cdot atndrte$。$priGPA$ 的系数就变成了在均值水平上的偏效应，而且我们还得到其标准误。（参见习题 6.14。）

6.3 拟合优度和回归元选择的进一步探讨

192 由于初学者都倾向于过分强调 R-平方，所以迄今为止，在我们评价回归模型时，对 R^2 大小的考虑还是不够的。如我们将看到的那样，基于 R-平方的大小而选择一组解释变量，可能会导致一些不合理的模型。我们在第 10 章将发现，从时间序列回归中得到的 R-平方可以人为地偏高，并导致有误导性的结论。

经典线性模型假定中没有要求 R^2 必须大于某个特定值；R^2 无非就是 y 的变异中有多少能用总体中的 x_1，x_2，…，x_k 解释。我们已经看到了几个 R-平方相当小的回归。这虽然意味着我们没有对影响 y 的几个因素进行解释，但并不意味着 u 中的因素与自变量相关。零条件均值假定 MLR.3 只是决定，我们是否得到了自变量在其他条件不变情况下之影响的无偏估计量，而 R-平方的大小与此则没有直接关系。

但要记住，在方程中增加变量时，R-平方的相对变化则十分有用：式 (4.41) 中检验联合显著性的 F 统计量，关键取决于不受约束模型和受约束模型的 R-平方之差。

调整 *R*-平方

大多数回归软件包都在报告 R-平方的同时，也报告一个被称为**调整 *R*-平方**（adjusted R-squared）的统计量。由于在多数应用研究中都报告调整 R-平方，而它又有某些有用的性质，所以我们在这一小节进行专门讨论。

为了看出如何调整通常的 R-平方，最好把它写成

$$R^2=1-\frac{\mathrm{SSR}/n}{\mathrm{SST}/n} \tag{6.20}$$

式中，SSR 为残差平方和；SST 为总平方和。与方程 (3.28) 相比，我们所做的只是将 SSR 和 SST 同时除以 n。这个表达式表明，R^2 实际上度量的是什么。定义 σ_y^2 为 y 的总体方差，并令 σ_u^2 表示误差项 u 的总体方差。（我们到目前为止，都是用 σ^2 来表示 σ_u^2，但在这里最好更明确一些。）**总体 *R*-平方**（population

R-squared）被定义为 $1-\sigma_u^2/\sigma_y^2$；这就是 y 的变异在总体中能用自变量解释的比例。这正是通常认为 R^2 所估计的东西。

我们知道，R^2 用 SSR/n 来估计 σ_u^2 是有偏误的。那么，为什么不用 SSR/($n-k-1$)来取代 SSR/n 呢？而且由于 SST/($n-1$)是 σ_y^2 的无偏估计量，所以我们还可以用 SST/($n-1$)来取代 SST/n。利用这些估计量，并由于 $\hat{\sigma}^2=$ SSR/($n-k-1$)，所以我们可以得到 R-平方：

$$\begin{aligned}\bar{R}^2 &\equiv 1-[SSR/(n-k-1)]/[SST/(n-1)]\\ &=1-\hat{\sigma}^2/[SST/(n-1)] \qquad (6.21)\end{aligned}$$

193 因为这个符号通常用来表示调整 R-平方，所以有时也称之为 **$\bar{R}$-平方**（R-bar squared）。

虽然调整 R-平方有时还被称为**修正 R-平方**，但由于这个名字意味着作为一个总体 R-平方的估计量，$\bar{R}^2$ 比 R^2 多少要好一些，但不幸的是，通常不会认为 $\bar{R}^2$ 是一个更好的估计量，所以这个名字并不妥当。虽然人们禁不住认为，$\bar{R}^2$ 纠正了 R^2 在估计总体 R-平方时的偏误，但事实却并非如此：两个无偏估计量之比不是一个无偏估计量。

$\bar{R}^2$ 的根本吸引力在于，它为在一个模型中另外增加自变量施加了惩罚。我们知道，在一个回归方程中增加一个新的自变量，不可能使 R^2 下降，这是因为，随着更多自变量的加入，SSR 不会上升（而通常都是下降）。但 $\bar{R}^2$ 的公式表明，它明显取决于自变量的个数 k。如果在回归中增加一个自变量，那么，虽然 SSR 下降，但回归中的 $df=n-k-1$ 也下降。于是，当回归中增加一个新的自变量时，SSR/($n-k-1$)可能上升，也可能下降。

代数上有如下有趣的现象：如果我们在回归方程中增加一个新的自变量，那么，当且仅当新变量的 t 统计量在绝对值上大于 1，$\bar{R}^2$ 才会有所提高。（对此的一个推广是，在回归中增加一组变量时，当且仅当这些新变量联合显著性的 F 统计量大于 1，$\bar{R}^2$ 才会有所提高。）于是，我们立即看到，使用 $\bar{R}^2$ 来决定一特定的自变量（或变量组）是否属于某个模型时，所得到的答案与标准的 t 或 F 检验不同（因为在传统的显著性水平上，大小为 1 的 t 或 F 统计量在统计上是不显著的）。

有时得到一个用 R^2 表示 $\bar{R}^2$ 的表达式也是很有用的。简单运算即给出

$$\bar{R}^2=1-(1-R^2)(n-1)/(n-k-1) \qquad (6.22)$$

比如，如果 $R^2=0.30$，$n=51$，$k=10$，那么 $\bar{R}^2=1-0.70(50)/40=0.125$。于是，对于很小的 n 和很大的 k，$\bar{R}^2$ 可能会远低于 R^2。实际上，如果通常的 R-平方很小，而且 $n-k-1$ 也很小，$\bar{R}^2$ 则确实有可能为负！比方说，你可以代入 $R^2=0.10$，$n=51$ 和 $k=10$ 来验证 $\bar{R}^2=-0.125$。$\bar{R}^2$ 为负表明，它相对自由度个数而言，是一个很差的拟合模型。

调整 R-平方有时和回归中通常的 R-平方一起报告，有时则只报告 $\bar{R}^2$，来取代 R^2。重要的是要记住，在式（4.41）中 F 统计量里用到的是 R^2，而不是 $\bar{R}^2$。同样，含有 $\bar{R}_r^2$ 和 $\bar{R}_{ur}^2$ 的 F 公式也是不成立的。

利用调整 *R*-平方在两个非嵌套模型中进行选择

在 4.5 节，我们知道了如何计算 F 统计量来检验一组变量的联合显著性；这就使我们能够决定，在一特定的显著性水平上，这一组变量中是否至少有一个会影响因变量。但这个检验不能让我们决定哪个变量有影响。在某些情形下，我们想选择一个没有多余自变量的模型，调整 R-平方刚好可以帮我们做到这一点。

在 4.4 节中的大型棒球俱乐部的薪水模型中，我们看到，*hrunsyr* 和 194 *rbisyr* 都不是个别显著的。但这两个变量高度相关，所以我们想在下面两个模型中选择一个：

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + u$$

和

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 rbisyr + u$$

由于这两个方程中没有哪一个是另一个的特殊情形，所以它们是**非嵌套模型**（nonnested models）。我们在第 4 章研究的 F 统计量只能让我们检验嵌套模型：一个模型（受约束模型）是另一个模型（不受约束模型）的一种特殊情形，参见方程（4.32）和方程（4.28）这两个受约束和不受约束模型的例子。一种可能的办法是，创造一个包含原来两个模型中所有解释变量的复合模型，然后利用 F 检验将原来的每个模型都针对这个一般化的复合模型而进行检验。这一做法的问题在于，这两个模型既可能都被拒绝，又可能没有一个被拒绝（正如 4.4 节中大型棒球俱乐部的薪水模型那样）。因此，它并非总能提供一种方法，用来区别含有非嵌套回归元的模型到底哪个好。

在棒球运动员的薪水回归中，包含 *hrunsyr* 的回归的 $\bar{R}^2$ 是 0.621 1，而包含 *rbisyr* 的回归的 $\bar{R}^2$ 是 0.622 6。于是，基于调整 R-平方来看，含有 *rbisyr* 的模型略微好一点。但其差别实际上相当小，从而我们在习题 4.16 中通过控制某些变量可能得到不同的答案。（因为两个非嵌套模型都包含五个参数，所以使用通常的 R-平方则得到同样的结论。）

当自变量组代表着不同的函数形式时，通过比较 $\bar{R}^2$ 而在不同的非嵌套自变量组之间进行选择也是有价值的。考虑两个将 R&D 与企业销售额相联系的模型：

$$rdintens = \beta_0 + \beta_1 \log(sales) + u \tag{6.23}$$

$$rdintens = \beta_0 + \beta_1 sales + \beta_2 sales^2 + u \tag{6.24}$$

第一个模型通过包括进来销售额的对数形式而刻画了收益递减的规律；第二个模型通过使用一个二次函数也刻画了收益递减的规律。于是，第二个模型比第一个模型多包含一个参数。

使用 RDCHEM.RAW 中对 32 家化工企业的观测数据，估计方程

(6.23) 时所得到的 R^2 是 0.061，而估计方程（6.24）时所得到的 R^2 是 0.148。因此，二次函数拟合得看起来要好得多。但由于第一个模型比式（6.24）包含的参数少一个，所以将通常的 R-平方进行比较，对第一个模型来说是不公平的。即式（6.23）是一个比式（6.24）更节省（parsimonious）的模型。

在所有其他条件相同的情况下，模型越简单越好。由于通常的 R-平方对更复杂的模型不加约束，所以最好是使用 $\bar{R}^2$ 来进行模型选择。式（6.23）的 $\bar{R}^2$ 是 0.030，而式（6.24）的是 0.090。于是，即使根据自由度的差异进行调整之后，二次函数模型也略胜一筹。在每个回归中都增加利润变元时，二次函数模型仍然更好一些。

195 在两个非嵌套模型之间进行选择时，利用 $\bar{R}^2$ 有一个重要的局限性：我们不能用它在因变量的不同函数形式之间进行选择。这很不幸，因为我们常常想基于拟合优度来决定，是应该用 y 作为因变量，还是应该用log(y)（或其他某种变形）作为因变量。但是，不论是 R^2 还是 $\bar{R}^2$，都不能用来决定这一点。原因很简单：不论我们在回归中使用的是什么因变量，R-平方所度量的都是因变量总变异中能被解释的比例。比如，y 和log(y)的总变异是不同的，将因变量形式不同的回归中所得到的调整 R-平方相比较，是不能在哪个模型拟合得更好这个问题上告诉我们任何信息的；它们拟合的是完全不同的两个因变量。

问题 6.4 ☞ 解释为什么通过最大化 R^2 或最小化 $\hat{\sigma}$（回归标准误）来选择模型是一回事。

例 6.4 CEO 的酬金与企业业绩

考虑将 CEO 酬金与企业业绩相联系的两个模型：

$$\widehat{salary} = 830.63 + 0.016\,3 sales + 19.63 roe \tag{6.25}$$
$$(223.90)\ (0.008\,9)\qquad (11.08)$$
$$n = 209,\ R^2 = 0.029,\ \bar{R}^2 = 0.020$$

和

$$\widehat{salary} = 4.36 + 0.275 lsales + 0.017\,9 roe \tag{6.26}$$
$$(0.29)\ (0.033)\qquad (0.004\,0)$$
$$n = 209,\ R^2 = 0.282,\ \bar{R}^2 = 0.275$$

式中，*roe* 为第 2 章讨论的股权回报，为简便起见，*lsalary* 和 *lsales* 分别表示 *salary* 和 *sales* 的自然对数。我们已经知道如何解释这两个不同的估计方程，但我们能说一个模型比另一个模型拟合得好吗？

方程（6.25）的 R-平方表明，*sales* 和 *roe* 只能解释样本中 CEO 薪水变异的约 2.9%。*sales* 和 *roe* 都具有边际的统计显著性。

方程（6.26）表明，log(*sales*)和 *roe* 解释了 log(*salary*)中的约 28.2%；用拟合优度的术语来说，看上去，这个高得多的 R-平方意味着模型（6.26）

好得多，但情况并非必然如此。样本中 salary 的总平方和是391 732 982，而 log(salary)的总平方和只有 66.72。因此，log(salary)中需要解释的变异要小得多。

从这一点来看，我们不妨使用 R^2 或 $\bar{R}^2$ 之外的一些特征，在这些模型之间作选择。比如，式（6.26）中的log(sales)和 roe 比式（6.25）中的 sales 和 roe 要统计显著得多，而式（6.26）中的系数可能更有意义。但为确定起见，我们还需要作出一个货真价实的拟合优度方面的比较。

196 在 6.4 节，我们将给出一个对拟合优度的度量，使得我们能在 y 以水平值形式和对数形式出现的模型之间进行比较。

回归分析中控制了过多的因素

在我们所讨论的许多例子中，当然在第 3 章中对遗漏变量偏误的讨论也不例外，我们都担心模型中会漏掉一个与自变量可能相关的重要因素。而还有另一种可能是，我们在一个回归分析中控制了过多的变量。如果我们过分强调拟合优度，我们就会在回归模型中无所顾忌地控制一些不应该控制的因素。为了避免这种错误，我们要记住对多元回归模型的其他条件不变的解释。

为说明这个问题，设想我们在进行一项研究，以评价州啤酒税对交通死亡率的影响。其思想是，较高的啤酒税将减少对酒精的消费，同样会减少醉酒驾车，从而导致较少的交通死亡率。为了度量税收在其他条件不变情况下对死亡率的影响，我们可以将 fatalities 模型化为包括啤酒税在内的几个因素的函数：

$$fatalities = \beta_0 + \beta_1 tax + \beta_2 miles + \beta_3 percmale + \beta_4 perc16-21 + \cdots$$

式中，miles 为驾驶的总里程数；percmale 为州人口中男性所占百分比；perc16－21 为人口中年龄在 16 岁－21 岁之间的人所占百分比，等等。注意我们为何没有包括一个能度量人均啤酒消费量的变量。我们在造成遗漏变量的偏误吗？答案是否定的。如果我们在这个方程中控制了啤酒消费量，那么，啤酒税会如何影响交通死亡率呢？在方程

$$fatalities = \beta_0 + \beta_1 tax + \beta_2 beercons + \cdots$$

中，β_1 度量了在保持 beercons 不变的情况下，死亡率因 tax 提高 1 个百分点而导致的差异。很难理解为什么这会有意义。除非我们想检验啤酒税的某种间接影响，否则就不应该控制 beercons 在州与州之间差别。诸如性别和年龄分布之类的其他因素，则应予以控制。

对某些特定因素该不该进行控制，并不总是清楚的。比如，贝茨（Betts，1995）所做的对高中学校的质量对以后挣钱的影响的研究。他指出，如果越好的学校质量将导致更多的教育，那么，在回归中在控制学校质量的

某种度量的同时又控制教育，将低估学校质量的回报。贝茨在方程中包括和不包括受教育年数的情况下分别进行了分析，以得到学校质量的估计影响的一个范围。

为了明显地看出关注高 R-平方怎么会引起麻烦，考虑 4.5 节中阐述的对多元假设进行检验的住房价格一例。其中，我们想检验住房价格评估的合理性。我们将$\log(price)$对$\log(assess)$，$\log(lotsize)$，$\log(sqrft)$和$bdrms$进行回归，并检验后三个变量是否具有零总体系数，并且$\log(assess)$有单位系数。但如果我们要像例 4.8 那样去估计一个享受价格模型，其中能得到各种住房特征的边际价值，结果会怎么样呢？我们是否应该在方程中包括$\log(assess)$呢？包括$\log(assess)$进行回归所得到的调整 R-平方为 0.762，而不包括它进行回归所得到的调整 R-平方为 0.630。仅从拟合优度来看，我们应该包括$\log(assess)$。但如果我们的目标是为了决定客厅大小、住房的平方英尺数和卧室间数对住房价格的影响，包括$\log(assess)$就不对了。方程中包括$\log(assess)$，等于是说保持住房价值的一种度量不变，然后问多一间卧室对住房价
197 值另一种度量有多大的影响。这对于评价住房特征来看将说不通。

如果我们记得不同的模型将起到不同的作用，并且注重对回归做其他条件不变的解释，我们就不会在一个回归模型中包括错误的因素。

增加回归元以减少误差方差

我们刚刚看过几个例子，其中有些自变量尽管与因变量相关，但也不应该包括在回归模型中。我们从第 3 章了解到，在回归中增加一个新的自变量会加剧多重共线性的问题。另一方面，由于从误差项中取出了一些因素作为解释变量，所以总可以减少误差方差，但一般而言，我们不知道哪方面的影响会占主导地位。

但有一种明显情形：我们总是应该把那些影响 y 而又与所有我们关心的自变量都无关的自变量包括进来。这样做的原因很简单：增加这样一个变量，不会导致总体出现多重共线性（因此样本中的多重共线性应该可以忽略），但却可以减小误差方差。在大样本容量的情况下，所有 OLS 估计量的标准误都将减小。

作为一个例子，考虑将个人对啤酒的需求作为全县平均啤酒价格的一个函数来估计。下面的假定也许是合理的：个人特征与全县水平的价格无关，所以将啤酒消费对全县啤酒价格的简单回归，足以估计价格对个人需求的影响。但通过包括诸如年龄和受教育程度等个人特征，可能会得到啤酒需求价格弹性的更精确的一个估计值。如果这些因素影响需求并与价格无关，那么，至少在大样本中，价格变量的标准误将较小。

不幸的是，在社会科学中，我们知道新增解释变量与所关心解释变量无关的情形相当少见。但值得记住的是，当我们能找到这些变量时，在不引起多重共线性的情况下，可以将它们包括到模型中，以减少误差方差。

6.4 预测和残差分析

我们在第3章定义了OLS预测值或拟合值和OLS残差。虽然预测确实有用，但由于预测值是通过使用OLS估计量而得到的，所以它们存在抽样波动的问题。现在，我们将在本节中说明，如何从OLS回归线得到一个预测值的置信区间。

从第3章和第4章我们知道，残差平方和及R-平方是通过残差得到的，所以残差对拟合优度和检验而言都很重要。经济学家有时候研究特定观测的残差，以了解样本中的个人（或企业、家庭等）。

预测的置信区间

198 假设我们有如下估计方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k \tag{6.27}$$

当我们将自变量的具体数值代入其中时，就得到y的一个预测值，它是在给定解释变量具体值情况下，对y的期望值的一个估计值。为强调起见，令c_1，c_2，…，c_k分别表示k个自变量中每一个自变量的具体值；这些具体值可能对应于样本中的某个实际的数据点，也可能没有哪个实际的数据点与之对应。我们想要估计的参数是

$$\theta_0 = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \cdots + \beta_k c_k = \mathrm{E}(y \mid x_1 = c_1, x_2 = c_2, \cdots, x_k = c_k) \tag{6.28}$$

θ_0的估计量是

$$\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \cdots + \hat{\beta}_k c_k \tag{6.29}$$

实际上，这很容易计算。但如果我们想得到对这个预测值的不确定性的某种度量，结果会怎么样呢？自然而然的办法是，以$\hat{\theta}_0$为中心，为θ_0构造一个置信区间。

为了得到θ_0的一个置信区间，我们需要$\hat{\theta}_0$的标准误。于是，在df较大的情况下，我们可以利用经验法则$\hat{\theta}_0 \pm 2 \cdot \mathrm{se}(\hat{\theta}_0)$来构造一个95%的置信区间。（和往常一样，我们也可以使用t分布中精确的百分位。）

怎么才能得到$\hat{\theta}_0$的标准误呢？这是在4.4节所遇到的同样的问题：需要得到OLS估计量之线性组合的一个标准误。由于所有的OLS估计量通常都出现在$\hat{\theta}_0$中（除非某些c_j为零），所以这里的问题要复杂得多。不过，在4.4节所用的方法在这里同样起作用。写出$\beta_0 - \theta_0 - \beta_1 c_1 - \beta_2 c_2 - \cdots - \beta_k x_k$，并代入方程

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k+u$$

就得到

$$y=\theta_0+\beta_1(x_1-c_1)+\beta_2(x_2-c_2)+\cdots+\beta_k(x_k-c_k)+u \tag{6.30}$$

换句话说，我们从对 x_j 的每一个观测中都减去 c_j 的值，然后

$$\text{将 } y_i \text{ 对 } (x_{i1}-c_1),(x_{i2}-c_2),\cdots,(x_{ik}-c_k),i=1,2,\cdots,n \text{ 进行回归} \tag{6.31}$$

从回归式（6.31）的截距（或常数项）就可以得到式（6.29）中的预测值，更重要的是，还能得到其标准误。

作为一个例子，我们为大学 *GPA* 回归的预测值得到一个置信区间，其中使用了高中学校的信息。

例 6.5 大学 *GPA* 预测值的置信区间

199 利用 GPA2.RAW 中的数据，我们得到预测大学 *GPA* 的如下方程：

$$\begin{aligned}\widehat{colgpa}=\ &1.493+0.001\,49sat-0.013\,86hsperc\\&(0.075)\quad(0.000\,07)\qquad(0.000\,56)\\&-0.060\,88hsize+0.005\,46hsize^2\\&(0.016\,50)\qquad\quad(0.002\,27)\\&n=413\,7,\ R^2=0.278,\ R^2=0.277,\ \hat{\sigma}=0.560\end{aligned} \tag{6.32}$$

我们在报告估计值时保留了几位数字，以减小四舍五入的误差。当 $sat=1\,200$, $hsperc=30$ 和 $hsize=5$（意味着 500）时，大学 *GPA* 的预测值是多少？通过将这些数值代入方程（6.32），很容易得到 $\widehat{colgpa}\approx 2.70$（保留两位小数）。不幸的是，我们不能用方程（6.32）直接得到 *colgpa* 的期望值在给定自变量值情况下的置信区间。得到置信区间的一个简单办法是，定义一个新的自变量组：$sat0=sat-1\,200$, $hsperc0=hsperc-30$, $hsize0=hsize-5$, $hsizesq0=hsize^2-25$。将 *colgpa* 对这些新的自变量进行回归时，得到

$$\begin{aligned}\widehat{colgpa}=\ &2.700+0.001\,49sat0-0.013\,86hsperc0\\&(0.020)\quad(0.000\,07)\qquad(0.000\,56)\\&-0.060\,88hsize0+0.005\,46hsizesq0\\&(0.016\,50)\qquad\quad(0.002\,27)\\&n=4\,137,\ R^2=0.278,\ \bar{R}^2=0.277,\ \hat{\sigma}=0.560\end{aligned}$$

此回归与式（6.32）中的回归相比，惟一的不同之处在于，它正是我们想预测的截距，其标准误是 0.020。斜率系数及其标准误与 *R*-平方等等都同前一样，这并非偶然；这提供了一种检查，即是否进行了适当的变换。我们轻而易举就能构造出预期大学 *GPA* 的一个 95% 的置信区间：$2.70\pm1.96(0.020)$ 或约 2.66～2.74。由于样本容量很大，所以这个置信区间相当窄。

因为当每个解释变量的样本均值都为零时，截距估计量的方差是最小的

（简单回归的情形，可参见问题2.5），所以从式（6.31）中的回归可得知，当 x_j 都取其均值时（即对所有的 j，都有 $c_j=\bar{x}_j$）预测值的方差是最小的。由于我们对接近数据中间值的回归线段最有信心，所以对这个结论不必过于吃惊。随着 c_j 的值越来越远离 $\bar{x}_j$，Var（$\hat{y}$）变得越来越大。

前面的方法，使我们对解释变量的任何值都能在 $E(y|x_1,x_2,\cdots,x_k)$ 的OLS估计值附近求得一个置信区间。但要得到 y 的一个新的（尚未知的）结果的置信区间，情况就不一样了。在对一个新结果构造一个置信区间时，我们必须考虑另一类十分重要的变异来源：观测不到的误差方差。

200 令 y^0 表示我们想为之构造一个置信区间——我们有时也称为**预测区间**（prediction interval）——的估计值。比如，y^0 可以表示不在我们原样本中的一个人或企业。令 x_1^0，…，x_k^0 为新的自变量值（假定我们能观测得到），并令 u^0 为观测不到的误差。因此有

$$y^0=\beta_0+\beta_1x_1^0+\beta_2x_2^0+\cdots+\beta_kx_k^0+u^0 \tag{6.33}$$

和前面一样，我们对 y^0 的最佳预测，就是给定解释变量，从OLS回归线估计 y^0 的期望值：$\hat{y}^0=\hat{\beta}_0+\hat{\beta}_1x_1^0+\hat{\beta}_2x_2^0+\cdots+\hat{\beta}_kx_k^0$。用来预测 y^0 的预测误差是

$$\hat{e}^0=y^0-\hat{y}^0=(\beta_0+\beta_1x_1^0+\beta_2x_2^0+\cdots+\beta_kx_k^0)+u^0-\hat{y}^0 \tag{6.34}$$

现在，由于 $\hat{\beta}_j$ 是无偏的，所以 $E(\hat{y}^0)=E(\hat{\beta}_0)+E(\hat{\beta}_1)x_1^0+E(\hat{\beta}_2)x_2^0+\cdots+E(\hat{\beta}_k)x_k^0=\beta_0+\beta_1x_1^0+\cdots+\beta_kx_k^0$。（和前面一样，这些期望值都以自变量的样本值为条件。）由于 u^0 的均值为零，所以 $E(\hat{e}^0)=0$。我们已经证明，预测误差的期望值为零。

在求 $\hat{e}^0$ 的方差时注意到，由于 u^0 与用来得到 $\hat{\beta}_j$ 的样本的方差不相关，所以 u^0 与每个 $\hat{\beta}_j$ 都不相关。根据协方差的基本性质（参见附录B），u^0 和 $\hat{y}^0$ 也不相关。于是，**预测误差的方差**（variance of the prediction error）就是二者方差之和：

$$\mathrm{Var}(\hat{e}^0)=\mathrm{Var}(\hat{y}^0)+\mathrm{Var}(u^0)=\mathrm{Var}(\hat{y}^0)+\sigma^2 \tag{6.35}$$

式中，$\sigma^2=\mathrm{Var}(u^0)$ 为误差方差。$\hat{e}^0$ 的方差有两个来源。第一个是 $\hat{y}^0$ 的抽样误差，来自我们对 β_j 的估计。由于每个 $\hat{\beta}_j$ 都有一个与 $1/n$ 成比例的方差，其中 n 是样本容量，所以 $\mathrm{Var}(\hat{y}^0)$ 与 $1/n$ 成比例。这意味着，对于大样本情形，Var（$\hat{y}^0$）可能会很小。对比之下，σ^2 则是总体误差的方差：它不随样本容量的变化而变化。在许多例子中，σ^2 都将是式（6.35）中占主导地位的一项。

在经典线性模型假定之下，$\hat{\beta}_j$ 和 u^0 都是正态分布的，所以 $\hat{e}^0$ 也是正态分布的（以解释变量的所有样本值为条件）。前面我们讨论过如何得到 $\mathrm{Var}(\hat{y}^0)$ 的一个无偏估计量，并在第3章得到了 σ^2 的一个无偏估计量。通过使用这些估计量，我们将 $\hat{e}^0$ 的标准误定义为

$$\mathrm{se}(\hat{e}^0)=\{[\mathrm{se}(\hat{y}_0)]^2+\hat{\sigma}^2\}^{1/2} \tag{6.36}$$

利用与 $\hat{\beta}_j$ 的 t 统计量同样的逻辑，$\hat{e}^0/\mathrm{se}(\hat{e}^0)$ 服从一个自由度为 $n-(k+1)$ 的 t 分布。于是

$$P[-t_{0.025} \leqslant \hat{e}^0/\mathrm{se}(\hat{e}^0) \leqslant t_{0.025}] = 0.95$$

式中，$t_{0.025}$ 为 t_{n-k-1} 分布中第 97.5 个百分位。对很大的 $n-k-1$，记住 $t_{0.025} \approx 1.96$。代入 $\hat{e}^0 = y^0 - \hat{y}^0$，经整理则给出 y^0 的一个 95% 的预测区间：

201

$$\hat{y}^0 \pm t_{0.025} \cdot \mathrm{se}(\hat{e}^0) \tag{6.37}$$

一如既往，除非 df 太小，否则，一个很好的经验法则就是 95% 的预测区间为 $\hat{y}^0 \pm 2\mathrm{se}(\hat{e}^0)$。由于式 (6.36)，所以这个区间比 $\hat{y}^0$ 本身的置信区间更宽；为了反映 u^0 中我们未加控制的因素，它常常要宽得多。

例 6.6 未来的大学 *GPA* 的置信区间

假使我们想得到一名高中学生未来的大学 *GPA* 的一个 95% CI，这名学生现在有 $sat = 1\,200$，$hsperc = 30$ 和 $hsize = 5$。记得我们在例 6.5 中得到了预期 *GPA* 的一个置信区间，现在我们必须解释误差项中的那些无法观测到的因素。我们已具备了得到 *colgpa* 的一个 CI 所需要的一切条件。$\mathrm{se}(\hat{y}^0) = 0.020$，$\hat{\sigma} = 0.560$，所以从式 (6.36) 可知，$\mathrm{se}(\hat{e}^0) = [(0.020)^2 + (0.560)^2]^{1/2} \approx 0.560$。注意，$\mathrm{se}(\hat{y}^0)$ 相对 $\hat{\sigma}$ 而言是多么小：几乎 $\hat{e}^0$ 中所有的变异都来自 u^0 的变异。95% 的 CI 为 $2.70 \pm 1.96(0.560)$ 或约为 1.60 - 3.80。这是一个很宽的置信区间，它表明，基于回归中所用到的因素，我们不可能显著地缩小大学 *GPA* 的可能范围。

残差分析

有时，检查一下个体观测值，看因变量的实际值是高于还是低于预测值也很有帮助；也就是考查个别观测的残差。这个过程被称为**残差分析**(residual analysis)。经济学家已经知道通过检查回归中的残差来帮助购买住房。下面这个住房价格的例子就说明了残差分析。住房价格与房子的各种可观测特征都有关系。我们可以列出我们所发现的所有重要特征，如大小、卧室间数等。我们可以使用住房的一个样本来估计价格与各种因素之间的关系，结果每一套住房都得到一个预测价值和一个实际价值。然后，就能构造出残差 $\hat{u}_i = y_i - \hat{y}_i$。残差负得越多的住房，至少基于我们所控制的因素而言，就是价格相对其特征被低估得最多的住房。使用方程 (6.37) 中描述的方法，为住房在未来可能的售价计算一个置信区间也讲得通。

利用 HPRICE1.RAW 中的数据，我们将 *price* 对 *lotsize*，*sqrft* 和 *bdrms* 做一个回归。在 88 套住房的样本中，负得最多的残差是第 81 套住房，为 −120.206。因此，这套住房的售价比其预测值要低 120 206 美元。

残差分析还有许多其他用处。对法学院进行排名的方法之一，是将起薪中位数对一系列学生特征（如新入学年级 *LAST* 分数的中位数、新入学年级大学 *GPA* 的中位数等）进行回归，并得到每个法学院的预测值和残差。残差最大的法学院预计具有最高的附加价值。（当然，一个人的起薪与法学院整体评价的中位数相比如何，仍有很大的不确定性。）这些残差可以与进入
202 每个法学院的成本一起使用来决定最好的价值，这将要求对未来的收益进行适当的贴现。

残差分析在司法决策中也能发挥作用。《纽约时报》（*New York Times*）中一篇题为《法官认为，是贫穷而不是隔离影响了学生的分数》的文章（6/28/95），描述了一个重要的司法案件。根本问题是，哈特福德校区在标准化考试中相对于周边郊区的表现不佳，是否因为学校高度隔离情况下落后的学校质量。法官断定："考试分数的差距，并不能说明哈特福德在教育学生上做得不够或很差，也不能说明其学校就是失败的，因为基于相关的社会经济因素而预测的分数正是人们预期的水平。"这个结论，几乎肯定是基于平均分数或分数的中位数对康涅狄格州各校区的社会经济特征的回归分析而得出的。法官的结论表明，给定哈特福德校区学生的贫穷水平，实际的考试分数类似于从回归分析中预测出来的分数：哈特福德的残差并没有负到一定程度，足以断定学校本身才是考试分数偏低的原因。

问题 6.5 你能否用残差分析来判断哪个电影演员得到相对票房而言过高的收入？

当因变量为 $\log(y)$ 时对 y 的预测

由于在经验研究中对因变量使用自然对数变换如此频繁，所以我们专门用这一小节来讨论因变量为 $\log(y)$ 时对 y 的预测。作为一个副产品，我们可以得到对对数模型拟合优度的一个度量，并能与水平值模型中的 R-平方相比。

为了得到预测值，定义 $\text{log}y=\log(y)$ 是有用的；这个记号用来强调，它是在如下模型中预测的 y 的对数：

$$\text{log}y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k+u \tag{6.38}$$

在此方程中，x_j 可能是其他变量的某些变换；比如，我们在 CEO 薪水的例子中，可以让 $x_1=\log(sales)$，$x_2=\log(mktval)$，$x_3=ceoten$。

给定 OLS 估计量，我们知道如何针对自变量的任意数值来预测 $\text{log}y$：

$$\widehat{\text{log}y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2+\cdots+\hat{\beta}_kx_k \tag{6.39}$$

现在，由于取对数函数的指数函数可将其还原成水平函数，所以我们首先猜想，预测 y 无非就是将 $\log(y)$ 的预测值换成指数函数值：$\hat{y}=\exp(\widehat{\text{log}y})$。但

这并不奏效：实际上，它将系统地低估 y 的预测值。因为，如果模型（6.38）服从 CLM 假定 MLR.1～MLR.6，那么就可以证明

$$E(y|\mathbf{x}) = \exp(\sigma^2/2)\cdot\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k)$$

式中，$\mathbf{x}$ 为自变量；σ^2 为 u 的方差。［如果 $u \sim \text{Normal}(0,\sigma^2)$，那么 $\exp(u)$ 的期望值就是 $\exp(\sigma^2/2)$。］这个方程表明，为了预测 y，需要进行一个简单的调整：

203 $$\hat{y} = \exp(\hat{\sigma}^2/2)\exp(\widehat{\log y}) \tag{6.40}$$

式中，$\hat{\sigma}^2$ 只不过是 σ^2 的无偏估计量。因为回归标准误 $\hat{\sigma}$ 总是要报告的，所以得到 y 的预测值就很容易。因为 $\hat{\sigma}^2>0$，所以 $\exp(\hat{\sigma}^2/2)>1$。对很大的 $\hat{\sigma}^2$，这个调整因子可能会显著地大于 1。

虽然式（6.40）中的预测不是无偏的，但它却是一致的。不存在 y 的无偏预测，而且在多数情况下式（6.40）就很好了，但它确实依赖于误差项 u 的正态性。我们在第 5 章证明了，即便 u 不服从正态分布，OLS 也具有令人满意的性质。所以，得到一个不依赖于正态性的预测很有帮助。若只假定 u 独立于解释变量，则我们就有

$$E(y|\mathbf{x}) = \alpha_0\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k) \tag{6.41}$$

式中，α_0 为 $\exp(u)$ 的期望值，并肯定大于 1。

给定一个估计值 $\hat{\alpha}_0$，就能将 y 预测为

$$\hat{y} = \hat{\alpha}_0\exp(\widehat{\log y}) \tag{6.42}$$

同样，它也只不过是要求将对数模型中的预测值求指数函数，并将结果乘以 $\hat{\alpha}_0$。

我们将看到，很容易就能得到的一个一致估计量。

当因变量为 log(y) 时对 y 的预测

（i）从 $\log y$ 对 x_1，x_2，…，x_k 的回归中得到拟合值 $\widehat{\log y_i}$。

（ii）对每个观测 i，都求出 $\hat{m}_i = \exp(\widehat{\log y_i})$。

（iii）现在，在不设截距的情况下求 y 对 $\hat{m}$ 的回归，即做一个过原点的简单回归。$\hat{m}$ 的系数（惟一的系数）就是 α_0 的估计值。

一旦得到了 $\hat{\alpha}_0$，就可以将它与 $\log y$ 的预测值一起来预测 y。步骤如下：

（i）对于给定的 x_1，x_2，…，x_k 的值，从式（6.39）求出 $\widehat{\log y}$。

（ii）利用式（6.42）直接得到预测值 $\hat{y}$。

例 6.7　对 CEO 薪水的预测

我们所考虑的模型是

$$\log(salary) = \beta_0 + \beta_1\log(sales) + \beta_2\log(mktval) + \beta_3 ceoten + u$$

所以 β_1 和 β_2 都表示弹性，而 $100\beta_3$ 则表示半弹性。利用 CEOSAL2.RAW

中的数据，估计的方程是

204

$$\begin{aligned}\widehat{lsalary} = {} & 4.504 + 0.163\,lsales + 0.109\,lmktval + 0.0117\,ceoten \\ & (0.257)\quad(0.039)\qquad\quad(0.050)\qquad\qquad(0.0053) \\ & n = 177,\ R^2 = 0.318\end{aligned} \tag{6.43}$$

式中，为清楚起见，令 *lsalary*，*lsales* 和 *lmktval* 分别表示 *salary*，*sales* 和 *mktval* 的对数。然后，对样本中的每一个观测都求出 $\hat{m}_i = \exp(\widehat{lsalary}_i)$。将 *salary* 对 $\hat{m}$ 进行回归（没有常数项）得出 $\check{\alpha}_0 \approx 1.117$。

我们可以将 $\check{\alpha}_0$ 的这个值与式（6.43）一起使用，根据 *sales*，*mktval* 和 *ceoten* 的任意数值来预测 *salary*。让我们求出 *sales* = 5 000（因为单位是百万美元，所以这里意味着是 50 亿美元），*mktval* = 10 000（或 100 亿美元）和 *ceoten* = 10 时薪水的预测值。从式(6.43)来看，*lsalary* 的预测值是 $4.504 + 0.163\cdot\log(5\,000) + 0.109\cdot\log(10\,000) + 0.0117(10) \approx 7.013$。因此预测的薪水是 $1.117\cdot\exp(7.013) \approx 1\,240.967$ 或 1 240 967 美元。如果忘记乘以 $\check{\alpha}_0 = 1.117$，所得到的预测就是 1 110 983 美元。

我们可以使用前面得到预测值的方法，来决定以 log*y* 作为因变量的模型在预测 *y* 时的表现怎么样。我们已经有了对 *y* 为因变量的模型的度量：*R*-平方和调整 *R*-平方。我们的目标是要找到 log*y* 模型中对拟合优度的一个度量，能与 *y* 为因变量的模型中的 *R*-平方相比。

虽然有几种方法都能找到这种度量，但我们还是给出一种易于实施的方法。在第（iii）步将 *y* 对 $\hat{m}$ 进行通过原点的回归之后，就得到这个回归的拟合值 $\check{y}_i = \check{\alpha}_0 \hat{m}_i$。然后我们求出样本中 $\check{y}_i$ 和实际的 y_i 之间的样本相关系数。其平方则可与我们通过在一个线性回归模型中用 *y* 作为因变量而得到的 *R*-平方相比。记住，拟合方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

中的 *R*-平方恰好就是 y_i 和 $\hat{y}_i$ 之相关系数的平方（参见 3.2 节）。

例 6.8　对 CEO 薪水的预测

在预测程序的第（iii）步之后，我们得到拟合值 $\check{salary}_i = \check{\alpha}_0 \hat{m}_i$。样本中 $salary_i$ 和 $\check{salary}_i$ 之间的简单相关系数是 0.493；其平方为 0.243。这就是我们对这个对数模型能在多大程度上解释 *salary* 的变异的一种度量；它不是式（6.43）中的 *R*-平方 0.318。

假设我们估计一个所有变量都以水平值出现的模型：

$$salary = \beta_0 + \beta_1 sales + \beta_2 mktval + \beta_3 ceoten + u$$

利用同样 177 个观测来估计这个模型所得到的 *R*-平方为 0.201。因此，对数模型能更多地解释 *salary* 中的变异，所以从拟合优度的角度来看，我们更

喜欢对数模型一些。对数模型还因为看起来更现实且其参数更易于解释而被选择

小　结

205　本章讨论了一些重要的多元回归分析的专题。

6.1节表明，改变自变量的度量单位，对OLS系数的改变恰如所料：如果x_j被乘以c，那么其系数就除以c。如果因变量被乘以c，那么所有的OLS系数都乘以c。改变任何一个变量的度量单位，都既不会影响t统计量，又不会影响F统计量。

我们讨论的β系数，以标准差为单位度量了自变量对因变量的影响。要得到β系数，首先将因变量和自变量都转换成z-得分，然后进行一个标准的OLS回归即可。

如我们在几个例子中所见到的那样，对数函数的形式赋予了系数百分比影响的解释。我们在6.2节又讨论了对数函数的其他几个优点。我们还看到，当一个对数—水平值模型的系数很大时如何计算准确的百分比影响。二次函数模型则容许出现边际影响递减或递增的情况。含有交互项的模型，则使得一个解释变量的边际影响取决于另一个解释变量的水平。

我们引入了调整R-平方$\bar{R}^2$，替代R^2作为度量拟合优度的另一种选择。尽管R^2在回归中增加另一个变量时不可能下降，但R^2则惩治了回归元的个数，并在增加自变量时可能下降。这就使得在解释变量个数不同的非嵌套模型之间进行选择时，$\bar{R}^2$更受欢迎一些。尽管R^2和$\bar{R}^2$都不能用于比较因变量不同的模型，但如在6.4节所见，当我们在y和$\log(y)$之间选择因变量时，要得到对拟合优度的可比较度量，则是相当容易的。

在6.3节，我们讨论了在到达最终模型的过程中会遇到的过度依赖R^2或$\bar{R}^2$这个多少有些微妙的问题：在一个回归模型中不可能控制太多的因素。出于这个原因，重要的是要预先考虑模型设定，特别是多元回归方程的其他条件不变之特性。既影响y又与所有其他解释变量无关的那些解释变量，可以在不导致多重共线性的情况下用来减小误差方差。我们在6.4节既说明了如何对从OLS回归线上得到的预测值构造置信区间，还说明了如何对y的未来的一个未知值构造置信区间。

我们偶尔还想在回归模型中的因变量是$\log(y)$时来预测y，6.4节就解释了这种方法。最后，我们有时想知道某些特殊观测的残差的符号和大小。可以使用残差分析来决定，某些特殊样本是否有远高于或远低于实际值的预测值之虞。

关键术语

调整 R-平方
β 系数
交互效应
206 预测误差
预测区间
二次函数
非嵌套模型
总体 R-平方
预测
残差分析
标准化系数
预测误差的方差

习 题

6.1 利用 CEOSAL1.RAW 中的数据估计了如下方程：

$$\log(\widehat{salary}) = 4.322 + 0.276\log(sales) + 0.021\,5\,roe - 0.000\,08\,roe^2$$
$$(0.324)\quad(0.033)\qquad\qquad(0.012\,9)\qquad(0.000\,26)$$
$$n = 209,\ R^2 = 0.282$$

这个方程使得 roe 对 $\log(salary)$ 具有边际递减的影响。这样概括“边际递减”是必然的吗？解释为什么是或为什么不是。

6.2 令 $\hat{\beta}_0$，$\hat{\beta}_1$，…，$\hat{\beta}_k$ 为 y_i 对 x_{i1}，…，x_{ik} 回归（$i = 1, 2, \cdots, n$）的 OLS 估计值。对于非零常数 c_1，…，c_k，证明 $c_0 y_i$ 对 $c_1 x_{i1}$，…，$c_k x_{ik}$ 回归（$i = 1, 2, \cdots, n$）的 OLS 截距和斜率由 $\tilde{\beta}_0 = c_0 \hat{\beta}_0$，$\tilde{\beta}_1 = (c_0/c_1)\hat{\beta}_1$，…，$\tilde{\beta}_k = (c_0/c_k)\hat{\beta}_k$ 给出。[提示：由于 $\hat{\beta}_j$ 是式（3.13）中一阶条件的解，从而 $\hat{\beta}_j$ 也必将是因变量和自变量重新测度后的一阶条件的解。]

6.3 使用 RDCHEM.RAW 中的数据，通过 OLS 得到如下方程：

$$\widehat{rdintens} = 2.613 + 0.000\,30\,sales - 0.000\,000\,007\,0\,sales^2$$
$$(0.429)\quad(0.000\,14)\qquad(0.000\,000\,003\,7)$$
$$n = 32,\ R^2 = 0.148\,4$$

(i) $sales$ 对 $rdintens$ 的边际影响在什么时候开始变成负的？

(ii) 你会在模型中保留二次项吗？请解释。

(iii) 定义 $sales$ 为以十亿美元计的销售额：$salesbil = sales/1\,000$。用 $salesbil$ 和 $salesbil^2$ 作为自变量重写估计方程。务必报告标准误和 R-平方。[提示：注意 $salesbil^2 = sales^2/(1\,000)^2$。]

(iv) 为了报告结果，你喜欢哪个方程？

6.4 如下模型使得受教育的回报还取决于父母双方受教育程度的总

和 *pareduc*：

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 educ \cdot pareduc + \beta_3 exper + \beta_4 tenure + u$$

(i)（以小数形式）证明此模型中多受一年教育的回报为

$$\Delta\log(wage)/\Delta educ = \beta_1 + \beta_2 pareduc$$

你预期 β_2 的符号是什么？为什么？

207 (ii) 利用 WAGE2.RAW 中的数据，所估计的方程是

$$\begin{aligned}
\log(\widehat{wage}) = {} & 5.65 + 0.047 educ + 0.000\,78 educ \cdot pareduc \\
& (0.13)(0.010) \qquad (0.000\,21) \\
& + 0.019 exper + 0.010 tenure \\
& (0.004) \qquad (0.003) \\
& n = 722,\ R^2 = 0.169
\end{aligned}$$

（只有 722 个观测包含了父母受教育的全部信息。）解释交互项的系数。选择 *pareduc* 的两个特殊值（比如，如果父母双方都受过大学教育，则 $pareduc = 32$，或父母都只受过高中教育，则 $pareduc = 24$），比较 *educ* 的估计回报。

(iii) 如果在方程中另增加一个变量 *pareduc*，则得到

$$\begin{aligned}
\log(\widehat{wage}) = {} & 4.94 + 0.097 educ + 0.033 pareduc \\
& (0.38)(0.027) \qquad (0.017) \\
& \quad 0.001\,6 educ \cdot pareduc - 0.020 exper + 0.010 tenure \\
& (0.001\,2) \qquad\qquad\qquad (0.004) \qquad (0.003) \\
& n = 722,\ R^2 = 0.174
\end{aligned}$$

现在，教育的估计回报正向地取决于父母的受教育水平吗？检验虚拟假设：教育的回报与父母的受教育水平无关。

6.5 在例 4.2 中，因变量是学生通过十年级数学考试（*math*10）的百分比，将 *sci*11（11 年级学生通过科学考试的百分比）作为另一个解释变量讲得通吗？

6.6 当我们把 $atndrte^2$ 和 $ACT \cdot atndrte$ 都增加到式（6.19）中的估计方程时，R-平方就变成 0.232。这些添加项在 10% 的显著性水平上是联合显著的吗？你会将它们包括在模型中吗？

6.7 如下三个方程是使用 401K.RAW 中的 1 534 个观测估计出来的：

$$\begin{aligned}
\widehat{prate} = {} & 80.29 + 5.44 mrate + 0.269 age - 0.000\,13 totemp \\
& (0.78)\ (0.52) \qquad (0.045) \qquad (0.000\,04) \\
& R^2 = 0.100, \bar{R}^2 = 0.098
\end{aligned}$$

$$\begin{aligned}
\widehat{prate} = {} & 97.32 + 5.02 mrate + 0.314 age - 2.66 \log(totemp) \\
& (1.95)\ (0.51) \qquad (0.044) \qquad (0.28) \\
& R^2 = 0.144, \bar{R}^2 = 0.142
\end{aligned}$$

$$\begin{aligned}
\widehat{prate} = {} & 80.62 + 5.34 mrate + 0.290 age - 0.000\,43 totemp \\
& (0.78)\ (0.52) \qquad (0.045) \qquad (0.000\,09)
\end{aligned}$$

$$+\,0.000\,000\,003\,9\,totemp^2$$
$$(0.000\,000\,000\,10)$$
$$R^2=0.108,\ \bar{R}^2=0.106$$

你更喜欢这三个模型中的哪一个？为什么？

计算机习题

208 6.8 仅使用 HPRICE3.RAW 中 1981 年的数据，回答如下问题。数据是 1981 年间在麻省 North Andover 售出住房的数据；1981 年是开始建造地方垃圾焚化炉的一年。

(i) 为了研究垃圾焚化炉的位置对住房价格的影响，考虑简单回归模型：

$$\log(price)=\beta_0+\beta_1\log(dist)+u$$

式中，*price* 为住房的美元价格；*dist* 为从住房到焚化炉的距离，以英尺为单位。谨慎地解释这个方程，如果焚化炉的出现会使住房价格下降，你预期 β_1 的符号是什么？估计这个方程，并解释你的结论。

(ii) 在第 (i) 部分的简单回归模型中增加变量 $\log(inst)$, $\log(area)$, $\log(land)$, *rooms*, *baths* 和 *age*，其中 *inst* 为从家到高速州际公路的距离；*area* 为住房的平方英尺数；*land* 为客厅的平方英尺数；*rooms* 为总的房间数；*baths* 为总的卫生间数；*age* 为住房的年数。现在，你对焚化炉的影响有什么结论？解释为什么第 (i) 部分和第 (ii) 部分给出相互矛盾的结论。

(iii) 向第 (ii) 部分的模型中添加 $[\log(inst)]^2$，结果会怎样？你对函数形式的重要性有什么结论？

(iv) 当你向第 (iii) 部分的模型中添加 $[\log(dist)]^2$ 时，它会显著吗？

6.9 本题利用 WAGE1.RAW 中的数据。

(i) 使用 OLS 估计方程

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 exper^2+u$$

并用通常的格式报告你的结论。

(ii) $exper^2$ 在 1% 的显著性水平上是统计显著的吗？

(iii) 使用近似

$$\%\Delta \widehat{wage}\approx 100(\hat{\beta}_2+2\hat{\beta}_3 exper)\Delta exper$$

求第 5 年工作经历的近似回报。第 20 年工作经历的近似回报是多少？

(iv) *exper* 取什么值时，工作经历的增加实际上会降低预期的 $\log(wage)$？样本中有多少人具有比这个取值更长的工作经历？

6.10 考虑一个教育回报取决于工作经历量（反之亦然）的模型：

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 educ\cdot exper+u$$

(i) 证明：保持 *exper* 不变，多受一年教育的回报（以小数表示）是 $\beta_1+\beta_3 exper$。

209 （ii）陈述如下虚拟假设：教育的回报并不取决于 $exper$ 的水平。你认为合适的对立假设是什么？

（iii）利用 WAGE2.RAW 中的数据，相对你给出的对立假设来检验（ii）中的虚拟假设。

（iv）令 θ_1 表示 $exper = 10$ 时（以小数表示）的教育回报：$\theta_1 = \beta_1 + 10\beta_3$。求出 θ_1 的估计值和一个95%的置信区间。[提示：写成 $\beta_1 = \theta_1 - 10\beta_3$ 并代入方程，然后重新整理，这就给出了得到 θ_1 的置信区间所需做的回归。]

6.11　本题利用 GPA2.RAW 中的数据。

（i）估计模型

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + u$$

式中，$hsize$ 为毕业年级的规模（以百为单位）。按通常的形式写出结论。二次项是统计显著的吗？

（ii）利用第（i）部分的估计方程，高中学校的“最优”规模是什么？说明你的答案。

（iii）这个分析是所有高中高年级学生学术成绩的代表吗？请解释。

（iv）用 $\log(sat)$ 作为因变量，求出估计的高中最优规模。它与你在第（ii）部分所得到的结论很不同吗？

6.12　本题利用 HPRICE1.RAW 中的数据。

（i）估计模型

$$\log(price) = \beta_0 + \beta_1\log(lotsize) + \beta_2\log(sqrft) + \beta_3 bdrms + u$$

并以通常的 OLS 格式报告结论。

（ii）当 $lotsize = 20\,000$，$sqrft = 2\,500$ 和 $bdrms = 4$ 时，求出 $\log(price)$ 的预测值。利用6.4节中的方法，在同样的解释变量值的情况下，求出 $price$ 的预测值。

（iii）就解释 $price$ 中的变异而言，决定你是喜欢第（i）部分的模型，还是喜欢模型

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

6.13　本题利用 VOTE1.RAW 中的数据。

（i）考虑一个含有竞选支出交互项的模型

$$\begin{aligned} voteA = {} & \beta_0 + \beta_1 prtystrA + \beta_2 expendA + \beta_3 expendB \\ & + \beta_4 expendA \cdot expendB + u \end{aligned}$$

保持 $prtystrA$ 和 $expendA$ 不变，$expendB$ 对 $voteA$ 的偏效应是什么？$expendA$ 对 $voteA$ 的偏效应是什么？β_4 的预期符号明显吗？

（ii）估计第（i）部分中的方程，并以通常的格式报告结果。交互项是统计显著的吗？

（iii）求样本中 $expendA$ 的均值。固定 $expendA$ 为300（300 000美元），

候选人 B 另外支出 100 000 美元对 *voteA* 的估计影响是什么？这个影响很大吗？

210 （iv）现在固定 *expendB* 为 100。$\Delta expendA = 100$ 对 *voteA* 的估计影响是什么？这讲得通吗？

（v）现在估计一个用候选人 A 的支出占竞选总支出的百分比 *shareA* 取代交互作用项的模型。同时保持 *expendA* 和 *expendB* 不变而改变 *shareA*，这讲得通吗？

（vi）（要求有微积分知识）在第（v）部分的模型中，保持 *prtystrA* 和 *expendA* 不变，求出 *expendB* 对 *voteA* 的偏效应。在 $expendA = 300$ 和 $expendB = 0$ 时进行计算，并评论你的结论。

6.14　本题利用 ATTEND.RAW 中的数据。

（i）在例 6.3 的模型中，推出

$$\Delta stndfnl/\Delta priGPA \approx \beta_2 + 2\beta_4 priGPA + \beta_6 atndrte$$

当 $priGPA = 2.59$ 和 $atndrte = 0.82$ 时，利用方程（6.19）来估计偏效应。对你的估计进行解释。

（ii）说明可将方程写成

$$stndfnl = \theta_0 + \beta_1 atndrte + \theta_2 priGPA + \beta_3 ACT + \beta_4 (priGPA - 2.59)^2 + \beta_5 ACT^2 + \beta_6 priGPA(atndrte - 0.82) + u$$

式中，$\theta_2 = \beta_2 + 2\beta_4(2.59) + \beta_6(0.82)$。（注意，截距已发生变化，但并不重要。）用它求出第（i）部分得到的 θ_2 的标准误。

6.15　本题利用 HPRICE1.RAW 中的数据。

（i）估计模型

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

并按通常的形式报告你的结果，包括回归的标准误。当我们代人 $lotsize = 10\,000$，$sqrft = 2\,300$ 和 $bdrms = 4$ 时，求出预测价格；将这个价格四舍五人到美元。

（ii）做一个回归，使你能得到第（*i*）部分中预测值的一个 95% 的置信区间。注意，由于四舍五人的误差，你的预测多少有些不同。

（iii）令 $price^0$ 为具有第（i）部分和第（ii）部分所述特征的住房未知的未来售价。求出 $price^0$ 的一个 95% 的置信区间，并对这个置信区间的宽度进行评论。

第 7 章　含有定性信息的多元回归分析：二值(或虚拟)变量

211 在前面几章中，我们的多元回归模型中的因变量和自变量都具有定量的含义。就像小时工资率、受教育年数、大学平均成绩、空气污染量、企业销售水平和被拘捕次数等。在每种情况下，变量的大小都传递了有用的信息。在经验研究中，我们还必须在回归模型中考虑定性因素。一个人的性别或种族、一个企业所属的产业（制造业、零售业等）和一个城市在美国所处的地理位置（南、北、西等）都可以被认为是定性因素。

本章的绝大部分内容都在探讨定性自变量。我们在 7.1 节介绍了描述定性信息之后，又在 7.2 节、7.3 节和 7.4 节中说明了如何在多元回归模型中很容易地包含定性的解释变量。这几节几乎涵盖了定性自变量用于横截面数据回归分析的所有流行方法。

我们在 7.5 节讨论了定性因变量的一种特殊情况，即二值因变量。这种情形下的多元回归模型具有一个有趣的含义，并被称为线性概率模型。尽管有些计量经济学家对线性概率模型多有中伤，但其简洁性还是使之在许多经验研究中有用武之地。虽然我们在 7.5 节将指出其缺陷，但在经验研究中，这些缺陷常常是次要的。

7.1 对定性信息的描述

定性信息通常以二值信息的形式出现：一个人是男还是女；一个人有还是没有一台个人计算机；一家企业向其一类特定的雇员提供还是不提供退休金方案；一个州实行或不实行死刑。在所有这些例子中，有关信息可通过定义一个**二值变量**（binary variable）或一个0—1变量来刻画。在计量经济学 212 中，对二值变量最常见的称呼是**虚拟变量**（dummy variable），尽管这个名称并不是特别形象。

问题 7.1 ☞ 假设在一项比较民主党和共和党候选人之间选举结果的研究中，你想标明每个候选人所在的党派。在这种情形中，名称 *party* 是二值变量的一个明智选择吗？更好的名称是什么？

在定义一个虚拟变量时，我们必须决定赋予哪个事件的值为1和哪个事件的值为0。比如，在一项对个人工资决定的研究中，我们可能定义 *female* 为一个虚拟变量，并对女性取值1，而对男性取值0。这种情形中的变量名称就是取值1的事件。通过定义 *male* 在一个人为男性时取值1并在一个人为女性时取值0，也能刻画同样的信息。这两种情况都比使用 *gender* 更好，因为这个名称没有指出虚拟变量何时取值1：*gender*=1对应于男性还是女性？虽然怎样称呼变量对得到回归结果而言并不重要，但它总有助于选择那些使方程和阐述都更清晰的变量。

表 7.1　　WAGE1.RAW 中的局部数据列表

个人编号	*wage*	*educ*	*exper*	*female*	*married*
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

假设我们在工资的例子中已选择了 *female* 来表示性别。此外，我们还定义了一个二值变量 *married*，并在一个人已婚时取值1，而在其他情况下

取值0。表7.1给出了可能得到的一个数据集的部分列表。我们看到，第一个人为女性并且未婚，第二个人为女性并且已婚，第三个人为男性并且未婚，等等。

我们为什么要用数值0和1来描述定性信息呢？在某种意义上，这些值是任意的：用任意两个不同的数值都是一样的。使用0—1变量来刻画定性信息的真正好处，像我们将看到的那样，在于它导致回归模型中的参数有十分自然的解释。

7.2 只有一个虚拟自变量

213 我们如何在回归模型中引入二值信息呢？在只有一个虚拟解释变量的最简单情形中，我们只在方程中增加一个虚拟变量作为自变量。比如，考虑如下决定小时工资的简单模型：

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + u \tag{7.1}$$

用 δ_0 表示 *female* 的参数，以强调虚拟变量参数的含义；以后，无论如何，我们还是使用最方便的符号。

在模型（7.1）中，只有两个被观测因素影响工资：性别和受教育水平。由于对女性 $female = 1$，而对男性 $female = 0$，所以参数 δ_0 具有如下含义：给定同等受教育程度（和同样的误差项 u），δ_0 是女性与男性之间在小时工资上的差异。因此，系数 δ_0 决定了是否对女人存在歧视：如果 $\delta_0 < 0$，那么在其他因素的相同水平下，女人总体上挣的钱要比男人少。

用期望的术语来讲，如果我们假定了零条件均值假定 $E(u \mid female, educ) = 0$，那么

$$\delta_0 = E(wage \mid female = 1, educ) - E(wage \mid female = 0, educ)$$

由于 $female = 1$ 对应于女性和 $female = 0$ 对应于男性，所以我们可以更简单地把这个模型写成

$$\delta_0 = E(wage \mid female, educ) - E(wage \mid male, educ) \tag{7.2}$$

这里的关键在于，在两个预期中，受教育水平是相同的；差值 δ_0 只是由于性别所致。

这种情况可以在图上描绘成男性与女性之间的**截距迁移**（intercept shift）。在图7.1中，给出了 $\delta_0 < 0$ 的情形，从而男人比女人每小时都多挣一个固定的数量。这个差距与受教育水平无关，这就解释了为什么女人和男人的工资—受教育变化关系是平行的。

这里，你可能想知道为什么我们没有在式（7.1）中包括一个虚拟变量 *male*，它对男性取值1和对女性取值0。原因在于，这样做是多余的。在式（7.1）中，男性线的截距是 β_0，女性线的截距是 $\beta_0 + \delta_0$。由于只有两组数据，

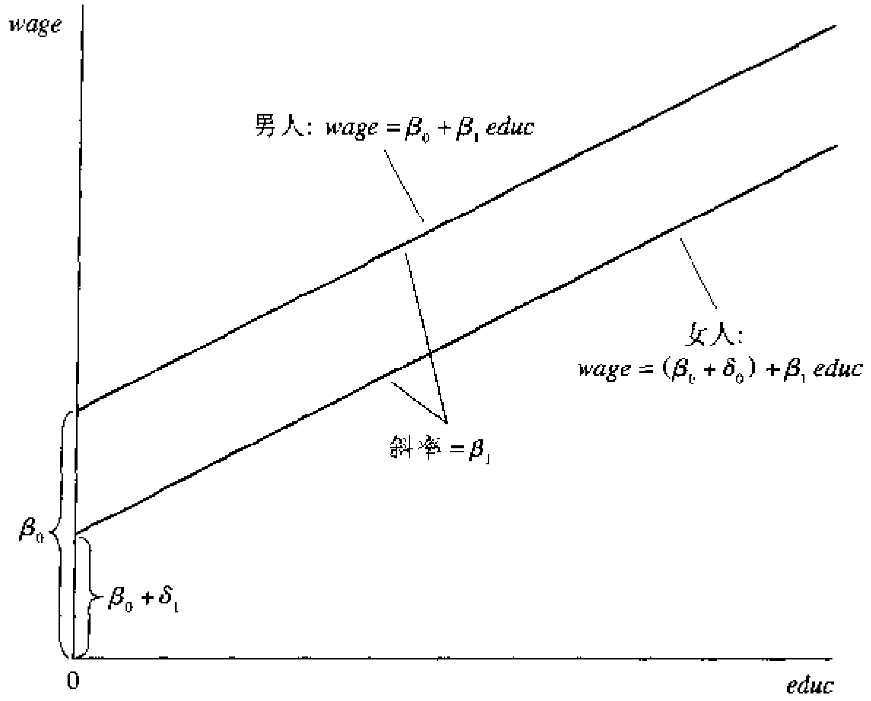


图 7.1　在 $wage = \beta_0 + \delta_0 female + \beta_1 educ + u$ 在 $\delta_0 < 0$ 情况下的图示

所以我们只需要两个不同的截距。这意味着，除了 β_0 之外，我们只需要一个虚拟变量；我们已经选择了针对女性的虚拟变量。由于 $female + male = 1$ 意味着 $male$ 是 $female$ 的一个完全线性函数，所以使用两个虚拟变量将导致完全多重共线性。包括两个性别的虚拟变量是所谓**虚拟变量陷阱**（dummy variable trap）中最简单的例子，当使用过多的虚拟变量来描述一定组数的数据时，就会掉进这种虚拟变量陷阱。我们以后还要讨论这个问题。在式（7.1）中，我们已经选择了男性为**基组**（base group）或**基准组**（benchmark
214 group），即与之进行比较的那一组。这就是为什么 β_0 表示了男性的截距，而 σ_0 为女性与男性之间在截距上的差异。通过将模型写成

$$wage = \alpha_0 + \gamma_0 male + \beta_1 educ + u$$

我们就能选择女性为基组，其中女性的截距是 α_0，而男性的截距是 $\alpha_0 + \gamma_0$；这意味着 $\alpha_0 = \beta_0 + \delta_0$ 和 $\alpha_0 + \gamma_0 = \beta_0$。在任何一个实际应用中，我们如何选择基组都不重要，但重要的是要保持基组不变。

有些研究者喜欢将模型中的总截距去掉，而将每一组的虚拟变量都包括进来。那么，这里的方程就是 $wage = \beta_0 male + \alpha_0 female + \beta_1 educ + u$，其中男人的截距是 β_0，女人的截距是 α_0。在这种情形下，因为没有总截距，所以不存在虚拟变量陷阱。但由于检验截距的差值更困难，而且对不含截距项的回归怎样计算 R-平方没有一个一致同意的方法，所以这个表达式很少有人使用。因此，我们将总是引进一个总的截距项作为基组的截距。

当解释变量更多时并没有什么大的改变。取男性那一组为基组，除了控
215 制受教育水平之外，还控制工作经历和现职任期的一个模型是

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u \qquad (7.3)$$

如果 *educ*，*exper* 和 *tenure* 都是相关的生产力特征，那么男人与女人之间没有差别的虚拟假设是 H_0：$\delta_0 = 0$。对立假设是，对女人存在歧视 H_1：$\delta_0 < 0$。

我们怎样才能对工资歧视进行实际检验呢？回答很简单：完全像从前那样用 OLS 来估计模型，并使用通常的 t 统计量。当某些自变量被定义为虚拟变量时，在 OLS 的操作和统计理论方面都没有任何改变。迄今为止，惟一的改变是我们对虚拟变量系数的解释。

例 7.1　小时工资方程

利用 WAGE1.RAW 中的数据，我们估计模型（7.3）。目前，我们还是使用 *wage* 而不是log(*wage*)作为因变量：

$$\begin{aligned}\widehat{wage} = {} & -1.57 - 1.81 female + 0.572 educ + 0.025 exper + 0.141 tenure \\ & (0.72)\quad (0.26) \qquad (0.049) \qquad (0.012) \qquad (0.021) \\ & n = 526, R^2 = 0.364 \end{aligned} \qquad (7.4)$$

负的截距（这里是男人组的截距）不是很有意义，因为样本中没有一个人具有接近于零年的 *educ*，*exper* 和 *tenure*。*female* 的系数则很有意思，因为它度量的是在给定相同水平的 *educ*，*exper* 和 *tenure* 的情况下，一个女人和一个男人之间每小时在工资上的平均差距。如果我们找到受教育水平、工作经历和现职任期相同的一个女人和一个男人，那么平均来看，女人每小时比男人要少挣 1.81 美元。（要记住，这可是用 1976 年的工资水平来度量的呀！）

重要的是记住，由于我们已经进行了多元回归并控制了 *educ*、*exper* 和 *tenure*，所以这 1.81 美元的工资差距不能由男人和女人之间在受教育水平、工作经历和现职任期水平上的平均差距来解释。我们可以断定，这 1.81 美元的差别，是由于性别或我们在回归中没有控制的与性别相关的因素所导致的。

将方程（7.4）中 *female* 的系数，与把所有其他解释变量都从方程中去掉时所得到的系数估计值相比，是颇有意义的：

$$\begin{aligned}\widehat{wage} = {} & 7.10 - 2.51 female \\ & (0.21)(0.30) \\ & n = 526,\ R^2 = 0.116 \end{aligned} \qquad (7.5)$$

216 式（7.5）中的系数具有一个简单的解释。这个截距就是样本中男人的平均工资（令 $female = 0$），所以男人平均每小时挣 7.10 美元。*female* 的系数为女人和男人之间平均工资的差距。因此，样本中女人的平均工资是 $7.10 - 2.51 = 4.59$，或每小时 4.59 美元。（顺便提一下，样本中有 274 个男人和 252 个女人。）

方程（7.5）为男人和女人这两组之间进行**均值比较检验**提供了一个简单的方法。估计差别 -2.51 的 t 统计量 -8.37 在统计上是十分显著的（当然 -2.51 在经济上也相当大）。一般而言，对一个常数和一个虚拟变量进行

简单回归，是比较两组均值的直接方法，要通常的 t 统计量生效，我们还必须假定同方差性，这就意味着，对男人和对女人而言，工资的总体方程是相同的。

由于方程（7.5）没有控制受教育水平、工作经历和现职任期上的差异，而且总体上说，在这个样本中，女人的受教育水平、工作经历和现职任期比男人要低些，所以估计出来的男女工资差异比方程（7.4）要大。方程（7.4）给出了性别工资差距在其他条件不变情况下更可靠的估计值，它仍显示出一个很大的差别。

在许多情形下，虚拟门变量都反映了个人或其他经济单位的选择（而不是诸如性别等预先决定的变量）。对于这种情形，因果关系的问题再度成为一个核心议题。在下面这个例子中，我们想知道，是否拥有个人计算机将导致更高的大学平均成绩。

例 7.2　拥有计算机对大学 *GPA* 的影响

为了决定拥有计算机对大学平均成绩的影响，我们估计了模型

$$colGPA = \beta_0 + \delta_0 PC + \beta_1 hsGPA + \beta_2 ACT + u$$

其中虚拟变量在学生拥有一台计算机时取值 1，在其他情况下都取值 0。出于各种原因，拥有 PC 对 *colGPA* 可能具有影响，一个学生的工作在计算机上完成的质量可能会更高一些，而且还因为不必在计算机房等待而节省了时间。当然，如果一个学生拥有一台 PC，他或她也可能会更热衷于电脑游戏或在网上冲浪，所以 δ_0 并不明显为正。变量 *hsGPA*（高中时的 *GPA*）和 *ACT*（能力测验分数）都是控制变量：能力越强的学生（用高中 *GPA* 和 *ACT* 分数来度量），越可能拥有一台计算机，控制这些因素是因为我们想知道，如果我们随机抽取一个学生并给他一台个人计算机，那么拥有这台计算机对其 *colGPA* 的平均影响。利用 GPA1.RAW 中的数据，我们得到

$$\begin{aligned}\widehat{colGPA} = {} & 1.26 + 0.157PC + 0.447hsGPA + 0.008\,7ACT \\ & (0.33)(0.057) \quad (0.094) \qquad (0.010\,5) \\ & n = 141, R^2 = 0.219\end{aligned} \tag{7.6}$$

217 这个方程意味着，一个拥有一台 PC 的学生，预计其 *GPA* 比一名条件相当但没有一台 PC 的学生要高出 0.16 分（记住，*colGPA* 和 *hsGPA* 都是以四分制度量的）。这个影响也是十分显著的，其 $t_{PC} = 0.157/0.57 \approx 2.75$。

如果我们从方程中去掉 *hsGPA* 和 *ACT* 会怎么样呢？显然，由于 *ACT* 的系数和 t 统计量都很小，所以去掉它几乎没有什么影响。但是 *hsGPA* 十分显著，所以去掉它会影响到 β_{PC} 的估计值。将 *colGPA* 对 *PC* 进行回归，给出 *PC* 的系数估计值约为 0.170，而标准误为 0.063；在此情形下，$\hat{\beta}_{PC}$ 及其 t 统计量不会改变太多。

在章末的习题中，要求你在方程中控制一些其他因素，以看拥有计算机的

影响会消失，或至少大大变小。

前面的每一个例子都可视为有**政策分析**（policy analysis）的含义。在第一个例子中，我们对工作中的性别歧视感兴趣。在第二个例子中，我们考虑了拥有计算机对大学成绩的影响。**项目评价**（program evaluation）是一种特殊的政策分析，我们在项目评价中想了解的是某些经济或社会项目对个人、企业、邻居、城市等的影响。

在最简单的项目评价中，把对象分为两组。**对照组**（control group）不参加这个项目，而**试验组**（experimental group）或**处理组**（treatment group）则参加。这些名称来自试验科学的文献，不应该从字面上进行理解。除非在极少数情形中，对对照组和处理组的选择都不是随机的。但在某些情况下，为了估计项目中的因果效应，可以使用多元回归分析来控制足够多的其他因素。

例 7.3　培训津贴对培训小时数的影响

利用 JTRAIN.RAW 中密歇根州制造企业在 1988 年的数据，我们得到如下估计方程：

$$\widehat{hrsemp} = \underset{(43.41)}{46.67} + \underset{(5.59)}{26.25}grant - \underset{(3.54)}{0.98}\log(sales) - \underset{(3.88)}{6.07}\log(employ) \qquad (7.7)$$

$$n = 105,\ R^2 = 0.237$$

因变量是在企业的水平上对每个雇员进行培训的平均小时数。变量 *grant* 是一个虚拟变量，1988 年得到在职培训津贴的企业取值 1，否则取值 0。变量 *sales* 和 *employ* 则分别表示了企业的年度销售额和雇员人数。因为变量 *hrsemp* 对回归中所用到的 105 个企业中的 29 个都取值 0，所以它不能以对数形式进入回归方程。

218 变量 *grant* 在统计是十分显著的，$t_{grant} = 4.70$。在控制了销售额和就业的情况下，得到津贴的企业平均对每个工人多培训 26.25 个小时。由于样本中平均每个工人培训的小时数约为 17，最大值也只有 164，所以 *grant* 像预期的那样对培训具有很大的影响。

log(*sales*)的系数很小而且极不显著。log(*employ*)的系数意味着，如果一个企业扩大 10%，那它对其工人的培训约减少 0.61 个小时。其 t 统计量为 −1.56，只是处在统计显著的边缘上。

就像所有其他的自变量一样，我们需要知道，对一个定性变量所度量的影响是不是因果性的。在方程（7.7）中，得到津贴和没有得到津贴的企业之间在培训上的差异，是由于津贴的原因吗？接受津贴会不会只是其他什么情况的一种表示呢？那些得到津贴的企业可能在没有得到津贴的情况下，一般也会更多地培训其工人。这个分析中并没有告诉我们所估计的是一种因果效应；我们必须知道，企业得到津贴是如何决定的。我们只能希望对于那些

与企业是否得到津贴及其培训水平相关的因素，我们已经尽可能多地加以控制。

在 7.6 节和以后的章节里，我们将回到用虚拟变量作政策分析。

当因变量为log(*y*)时，对虚拟解释变量系数的解释

在应用研究中有一个常见的设定，当自变量中有一个或多个虚拟变量时，因变量则以对数形式出现。在这种情况下，我们该如何解释虚拟变量的系数呢？无足为奇，此系数具有一种**百分比解释**。

例 7.4　住房价格回归

利用 HPRICE1.RAW 中的数据，我们得到方程

$$\begin{aligned}\log(\widehat{price}) = {} & \underset{(0.65)}{5.56} + \underset{(0.038)}{0.168}\log(lotsize) + \underset{(0.093)}{0.707}\log(sqrft) \\ & + \underset{(0.029)}{0.027}bdrms + \underset{(0.045)}{0.054}colonial \\ & n = 88,\ R^2 = 0.649\end{aligned} \tag{7.8}$$

除了二值变量 *colonial* 之外，所有的变量都无须多加解释。如果住房是殖民地建筑风格的，则 $colonial = 1$。*colonial* 的系数有什么含义呢？对于给定的 *lotsize*，*sqrft* 和 *bdrms* 的水平，一套殖民地建筑风格的住房与其他风格的住房在$\log(\widehat{price})$上的差别是 0.054。这意味着，在保持其他因素不变的情况下，一套殖民地建筑风格的住房的卖价预计约高出 5.4%。

219 这个例子表明，当log(*y*)是一个模型的因变量时，将虚拟变量的系数乘上 100，可解释为 *y* 在保持所有其他因素不变的情况下的百分比差异。当一个虚拟变量的系数表现出 *y* 较大比例的变化时，那么，完全像 6.2 节中对半弹性的计算一样，可以得到精确的百分比差异。

例 7.5　对数小时工资方程

让我们将例 7.1 中工资方程的因变量换成log(*wage*)，并增加 *exper* 和 *tenure* 的二次项，来重新估计它：

$$\begin{aligned}\log(\widehat{wage}) = {} & \underset{(0.099)}{0.417} - \underset{(0.036)}{0.297}female + \underset{(0.007)}{0.080}educ + \underset{(0.005)}{0.029}exper \\ & - \underset{(0.000\,10)}{0.000\,58}exper^2 + \underset{(0.007)}{0.032}tenure - \underset{(0.000\,23)}{0.000\,59}tenure^2 \\ & n = 526,\ R^2 = 0.441\end{aligned} \tag{7.9}$$

利用例 7.4 中同样的近似，*female* 的系数意味着，在 *educ*，*exper* 和 *tenure*

的相同水平上，女人比男人约少挣100(0.297) = 29.7%。通过计算预期工资上精确的百分比差异，我们可以做得比这更好。我们想得到的是，在保持所有其他因素都不变的情况下，女性与男性工资差异的比例：$(\widehat{wage}_F - \widehat{wage}_M)/\widehat{wage}_M$。从式（7.9），我们得到

$$\log(\widehat{wage}_F) - \log(\widehat{wage}_M) = -0.297$$

将它求指数函数并减去 1，则得到

$$(\widehat{wage}_F - \widehat{wage}_M)/\widehat{wage}_M = \exp(-0.297) - 1 \approx -0.257$$

这个更准确的估计值意味着一个女人的工资比一个与她相当的男人的工资大致低 25.7%。

如果我们在例 7.4 中进行同样的修正，将得到 exp（0.054）－1≈0.055 5，或约 5.6%。在例 7.4 中的修正比在工资一例中具有较小的影响，因为式（7.8）虚拟变量系数的大小比式（7.9）中虚拟变量的系数要小得多。

一般地，如果 $\hat{\beta}_1$ 是一个虚拟变量（比方说 x_1）的系数，那么，当 log(y) 是因变量时，在 $x_1 = 1$ 时预测的 y 相对于在 $x_1 = 0$ 时预测的 y，精确的百分比差异为

$$100\cdot[\exp(\hat{\beta}_1) - 1] \tag{7.10}$$

估计值 $\hat{\beta}_1$ 可正可负，重要的是，在计算式（7.10）时保留了它的符号。

7.3 使用多个虚拟变量

220 我们可以在同一个方程中使用几个虚拟自变量。比如，我们可以在方程（7.9）中增加一个虚拟变量 *married*。*married* 的系数给出了在保持性别、*educ*，*exper* 和 *tenure* 不变的情况下，那些已婚和未婚的人在工资上（近似）的比例差异。当我们估计这个模型时，*married* 的系数（标准误放在括号中）为 0.053（0.041），*female* 的系数则变成－0.290（0.036）。于是，估计“婚姻加薪”约为 5.3%，但在统计上并不显著异于零（$t = 1.29$）。这个模型的一个重要局限在于，假定了婚姻加薪对男人和女人而言都是一样的；下面这个例子则放松了这个假定。

例 7.6　对数小时工资方程

让我们估计一个工资对如下四组人都不同的模型：已婚男人、已婚女人、单身男人和单身女人。为了进行估计，必须选择一个基组；我们选择单身男人组。于是，我们必须对剩下的每一组都定义一个虚拟变量，并称之为

marrmale、*marrfem* 和 *singfem*。将这些变量代入式（7.9）（当然要去掉现在多余的变量 *female*），则给出

$$
\begin{aligned}
\log(\widehat{wage}) = {} & 0.321 + 0.213\,marrmale - 0.198\,marrfem \\
& (0.100)(0.055) \qquad\qquad (0.058) \\
& - 0.110\,singfem + 0.079\,educ \\
& (0.056) \qquad\qquad (0.007) \\
& + 0.027\,exper - 0.000\,54\,exper^2 + 0.029\,tenure \\
& (0.005) \qquad\quad (0.000\,11) \qquad\quad (0.007) \\
& - 0.000\,53\,tenure^2 \\
& (0.000\,23) \\
& n = 526,\ R^2 = 0.461
\end{aligned}
\tag{7.11}
$$

除 *singfem* 外，所有系数的 t 统计量在绝对值上都远大于 2；*singfem* 的 t 统计量约为 −1.96，相对双侧对立假设，刚刚在 5% 的显著性水平上显著。

为了解释虚拟变量的系数，我们必须记得，我们选择了单身男人组为基组。因此，三个虚拟变量的估计值度量的都是与单身男人相比，有工资的比例差异。比方说，在保持受教育水平、工作经历和现职任期不变的情况下，已婚男人约比单身男人多挣 21.3%。［式（7.10）中更精确的估计值约为 23.7%。］另一方面，在其他变量相同的情况下，预计一个已婚女人比一个单身男人少挣 19.8%。

由于基组用式（7.11）中的截距表示，所以我们只包括了四组中的三个作为虚拟变量。如果要在式（7.11）中增加一个单身男人的虚拟变量，那将因导致完全共线性而陷入虚拟变量陷阱。某些回归软件包将自动为你修正这个错误，而其他一些软件包则只告诉你存在着完全共线性。最好是细心地设
221 定虚拟变量，因为它能使我们正确地解释最终的模型。

尽管单身男人组是式（7.11）中的基组，但我们还是可以用这个方程来得到任意两组之间的估计差异。由于总体上的截距对每一组都是相同的，所以我们在找出差异时可以忽略它。因此，估计单身女人和已婚女人的差异约为 $-0.110-(-0.198)=0.088$，这意味着单身女人比已婚女人约多挣 8.8%。不幸的是，我们不能利用方程（7.11）来检验单身女人和已婚女人之间的估计差异是否统计显著。仅知道 *marrfem* 和 *singfem* 的标准误还不足以进行这个检验（参见 4.4 节）。最容易做到的是选择二者之一作为基组，并重新估计这个方程。虽然没有什么明显的变化，但直接得到了我们所需要的估计值及其标准误。当用已婚女人组作为基组而重新估计时，我们得到

$$
\begin{aligned}
\log(\widehat{wage}) = {} & 0.123 + 0.411\,marrmale + 0.198\,singfem \\
& (0.106)(0.056) \qquad\qquad (0.058) \\
& + 0.088\,singfem + \cdots \\
& (0.052)
\end{aligned}
$$

当然，其中未报告的系数或标准误都没有变化。恰如所料，*singfem* 的估计值为 0.088。现在，我们在得到这个估计值的同时也得到一个标准误。对于

总体中已婚女人和单身女人的工资没有差异的虚拟假设，t 统计量为 $t_{singfem}=0.088/0.052\approx1.69$。这只是拒绝虚拟假设的微弱证据。我们还看到，已婚男人和已婚女人的估计差异在统计上是十分显著的（$t_{marrmale}=7.34$）。

问题 7.2　在 MLB1.RAW 中发现的棒球运动员薪水的数据中，运动员有如下六个位置可供选择：*frstbase*、*scndbase*、*thrdbase*、*shrtstop*、*outfield* 和 *catcher*。为了说明不同位置上薪水的差异，我们以外场手（outfield）那一组为基组，你将把哪些虚拟变量作为自变量？

前面这个例子说明了在方程中包括虚拟变量来象征不同组的一般原则：如果回归模型具有 g 组或 g 类的不同截距，那我们就需要在模型中包含 $g-1$ 个虚拟变量和一个截距。基组的截距就是总体上的截距，某一组的虚拟变量的系数，则表示了该组与基组之间在截距上的估计差异。包括 g 个虚拟变量和一个截距，将导致虚拟变量陷阱。另一种办法是包括 g 个虚拟变量而没有总体截距，但如果这样的话，检验相对基组的差异就变得困难，所以这样做不够明智，某些回归软件包则在回归中没有包含截距项时改变计算 R-平方的方法。

通过虚拟变量来包含序数信息

假设我们想估计城市信用等级对市政债券利率（MBR）的影响。穆迪投资服务公司和标准普尔等几家金融公司，对地方政府债券的质量进行了级
222 别评定，其等级取决于违约概率等因素。（地方政府为降低其融资成本而喜欢较低的利率。）为简便起见，假设等级的范围是从零到四，零为最低的信用等级，四为最高的信用等级。这就是一个**序数变量**（ordinal variable）的例子。为简便起见，称这个变量为 CR。我们需要提出的问题是：如何将变量 CR 放到一个模型中去解释 MBR 呢？

一种可能是，就像包括所有其他解释变量一样把它包括进来：

$$MBR=\beta_0+\beta_1 CR+\text{其他因素}$$

我们没有明确说明模型中其他因素指的是什么。那么，β_1 就是保持其他因素不变，当 CR 增加一个单位时 MBR 的百分比变化。不幸的是，很难解释 CR 一个单位的变化。我们知道多一年受教育水平或每个学生多花 1 美元所包含的数量信息，但像信用等级之类的变量，典型地只有序数上的含义。我们知道 CR 为四比 CR 为三更好，但四级与三级之间的差距与一级和零级之间的差距一样吗？如果不一样，假定 CR 提高一个单位对 MBR 的影响为一个常数就讲不通。

由于 CR 只取相当少的几个数值，所以我们能使用的一个更好的方法是

对 CR 的每个值都定义一个虚拟变量。因此，如果 $CR=1$，则 $CR_1=1$，否则 $CR_1=0$；如果 $CR=2$，则 $CR_2=1$，否则 $CR_2=0$；如此等等。实质上，我们把信用等级分为五个类别。然后，我们可以估计模型

$$MBR=\beta_0+\delta_1 CR_1+\delta_2 CR_2+\delta_3 CR_3+\delta_4 CR_4+\text{其他因素} \qquad (7.12)$$

问题 7.3☞ 在模型（7.12）中，你如何检验信用等级对 MBR 没有影响的虚拟假设？

根据我们在模型中包括虚拟变量的规则，由于有五个类别，所以包括四个虚拟变量。这里省掉的一类是等于零的信用等级，所以它就是基组。（这就是为什么我们不需要对这一类别定义一个虚拟变量。）系数都很容易解释：信用等级为一级的城市和信用等级为零级的城市之间在 MBR 上的差异（保持其他因素不变）；为信用等级为二级的城市与信用等级为零级的城市之间在 MBR 上的差异；如此等等。因为这里使得每两个信用等级之间的变动都可能具有不同的影响，所以使用式（7.12）比简单地将 CR 作为一个单独变量代入方程更灵活。一旦定义了虚拟变量，估计则是相当容易的。

例 7.7 相貌吸引力对工资的影响

哈默什和比德尔（Hamermesh and Biddle，1994）在一个工资方程中使用了对相貌吸引力的某种度量。样本中的每一个人，都被面试主考官根据相貌的吸引力而归为五类（不好看、相当普通、一般水平、好看、特别漂亮或潇洒）中的某一类。因为很少有人处在两个极端上，所以作者将人分为三类进行回归分析：一般水平、低于一般水平和高于一般水平，其中一般化的那 223 一组是基组。利用来自 1977 年就业质量调查中的数据，在控制了通常的生产力特征之后，哈默什和比德尔对男人估计了方程：

$$\log(wage)=\hat{\beta}_0-0.164\,belavg+0.016\,abvavg+\text{其他因素}$$
$$(0.046)\qquad\qquad(0.033)$$
$$n=700,R^2-0.403$$

并对女人估计了方程：

$$\log(wage)=\hat{\beta}_0-0.124\,belavg+0.035\,abvavg+\text{其他因素}$$
$$(0.066)\qquad\qquad(0.049)$$
$$n=409,\bar{R}^2=0.330$$

回归中控制的其他因素包括受教育水平、工作经历、终身待遇、婚姻状况和种族等；对于更详尽的罗列，参见哈默什和比德尔文章中的表 3。为节省篇幅，文章中未报告其他变量的系数和截距。

对于男人，那些相貌低于平均水平的人，在其他方面相同（包括受教育水平、工作经历、终身待遇、婚姻状况和种族等）的情况下，预计比相貌处在平均水平的男人约少挣 16.4%。这个影响在统计上显著异于零，t 统计量为 -3.57。类似地，相貌高于平均水平的男人预计要多挣约 1.6%，尽管这

种影响在统计上并不显著（$t<0.5$）。

一个相貌低于平均水平的女人，比一个其他方面相当但相貌处在平均水平的女人约少挣12.4%，t 统计量为 −1.88。与男人的情况一样，*abvravg* 的估计值在统计上并不显著异于零。

在某些情况下，序数变量取值过多，以致不能对每个值都包括进来一个虚拟变量。比如，文件 LAWSCH85.RAW 包含了法学院毕业生起薪中位数的数据。一个关键的解释变量是法学院的排名。由于每个法学院都有一个排名，所以显然不能对每个排名都包括进来一个虚拟变量。如果我们不想直接把排名放到方程中，那就可以把它分成几类。下面这个例子就说明了这种做法。

例 7.8　法学院排名对起薪的影响

定义虚拟变量 $top10$，$r11\sim25$，$r26\sim40$，$r41\sim60$，$r61\sim100$，并让这些变量在排名落在相应的区间时取值 1。我们以排名在 100 名以后的法学院为基组，所估计的方程是

224

$$\begin{aligned}\log(\widehat{salary}) = {} & \underset{(0.41)}{9.17} + \underset{(0.053)}{0.700}top10 + \underset{(0.039)}{0.594}r11\sim25 + \underset{(0.034)}{0.375}r26\sim40 \\ & + \underset{(0.028)}{0.263}r41\sim60 + \underset{(0.021)}{0.132}r61\sim100 + \underset{(0.0031)}{0.0057}LSAT \\ & + \underset{(0.074)}{0.014}GPA + \underset{(0.026)}{0.036}\log(libvol) + \underset{(0.0251)}{0.0008}\log(cost) \qquad (7.13) \\ & n = 136, R^2 = 0.911, \bar{R}^2 = 0.905\end{aligned}$$

我们立即看到，所有根据不同排名定义的虚拟变量在统计上都十分显著。$r61\sim100$ 的估计值意味着，在保持 $LSAT$，GPA，$libvol$ 和 $cost$ 不变的情况下，排名在 61 到 100 之间的法学院的毕业生，与排名在 100 之后的法学院的毕业生相比，起薪的中位数要高约 13.2%。前 10 名的法学院与 100 名之后的法学院之间的差别就相当大了。使用方程（7.10）给出的精确计算，得到 $\exp(0.700)-1\approx1.014$，所以，预计前 10 名法学院毕业生的起薪中位数比 100 名之后法学院毕业生的起薪中位数要高出 100% 以上。

将排名分成不同的组是否标志着一种改进呢？不妨将式（7.13）中的调整 R-平方与把排名作为一个单独变量时得到的调整 R-平方相比较：前者是 0.905，而后者是 0.836，所以在式（7.13）中增加了回归的灵活性。

有意思的是，一旦将排名放到（无可否认，多少有些随意）给定的分类中，所有其他的变量都变得不显著了。实际上，对 $LSAT$，GPA，$\log(libvol)$ 和 $\log(cost)$ 联合显著性的检验给出的 p 值为 0.055，介于显著与不显著之间，当 $rank$ 以其原有形式被包括在模型中时，联合显著性检验的 p 值在小数点后四位小数都是零。

对此例的最后一点的评论：在推导普通最小二乘性质的过程中，我们假定了使用的是随机样本。在本例中，一个学院的排名必然取决于样本中其他学院的排名，所以数据不能说是从所有法学院中独立抽取的，这就违背了上述假定。但由于误差项与解释变量不相关，所以不会导致任何严重问题。

7.4 涉及虚拟变量的交互作用

虚拟变量之间的交互作用

就像具有定量意义的变量在回归模型中可以交互作用一样，虚拟变量也能产生交互作用。在例 7.6 中我们其实已经看到了这样的一个例子，其中我们根据婚姻状况和性别定义了四个类别。事实上，我们可以在 *female* 和 *married* 分别出现的模型中，增加一个 *female* 和 *married* 的**交互项**（interaction term）而重建这个模型。这就使得婚姻对薪金的升水就像在方程（7.11）中那样与性别有关。为便于比较，所估计的含有 *female* - *married* 交互项的模型为

225

$$\begin{aligned} \log(wage) = {} & 0.321 \quad 0.110 female + 0.213 married \\ & (0.100)(0.056) \qquad (0.055) \\ & - 0.301 female \cdot married + \cdots \\ & (0.072) \end{aligned} \tag{7.14}$$

其中回归的其余部分必定与式（7.11）一样。方程（7.14）明确表明，性别和婚姻状况之间存在着统计显著的交互作用。这个模型还使我们能得到所有四组之间的预期工资差异，但这里必须小心地将 0 和 1 的组合代入。

取 $female=0$ 和 $married=0$，这就排除了 *female*，*married* 和 $female \cdot married$，所以对应于单身男人这个基组。通过在式（7.14）中取 $female=0$ 和 $married=1$，就给出已婚男人组的截距，这个截距是 $0.321+0.213=0.534$，等等。

方程（7.14）不过是得到各种性别—婚姻状况组合之间工资差异的一种不同方法，它并不比方程（7.11）有什么真正的优势；实际上，方程（7.11）在检验任何一组与作为基组的单身男人组之间的差异时更容易些。

例 7.9　计算机使用对工资的影响

克鲁格（Krueger，1993）估计了计算机使用对工资的影响。他定义了一个被称为 *compwork* 的虚拟变量，此变量在一个人工作中使用了计算机时取值 1。另一个虚拟变量 *comphome* 则在一个人在家使用计算机时取值 1。

利用1989年人口普查中13 379个人的样本，克鲁格（Krueger，1993，表4）得到

$$\begin{aligned}\log(wage) = \hat{\beta}_0 &+ \underset{(0.009)}{0.177}compwork + \underset{(0.019)}{0.070}comphome \\ &+ \underset{(0.023)}{0.017}compwork \cdot comphome + \text{其他因素}\end{aligned} \qquad (7.15)$$

（其他因素就是工资回归中的标准因素，包括受教育水平、工作经历、性别和婚姻状况等；准确的列表可参见克鲁格的论文。）克鲁格没有报告截距，因为它没有任何重要性；我们所需要知道的一切，就是由那些在工作中和在家都不使用计算机的人构成的基组。值得注意的是，在工作中使用计算机（但在家里不使用）者的估计回报约高出17.7%，（更精确的估计值是19.4%）类似地，一个在家里使用计算机但在工作中不使用的人，与那些根本就不使用计算机的人相比，工资约高出7%。在两种情况下都使用计算机的人，比那些在两种情况下都不使用计算机的人，工资约高出26.4%（通过将三个系数相加并乘以100而得到），从方程（7.10）得到这种工资差距更精确的估计值为30.2%。

式（7.15）中的交互项在统计上不显著，在经济上也不是很大，但把它放在方程中也没有带来什么害处。

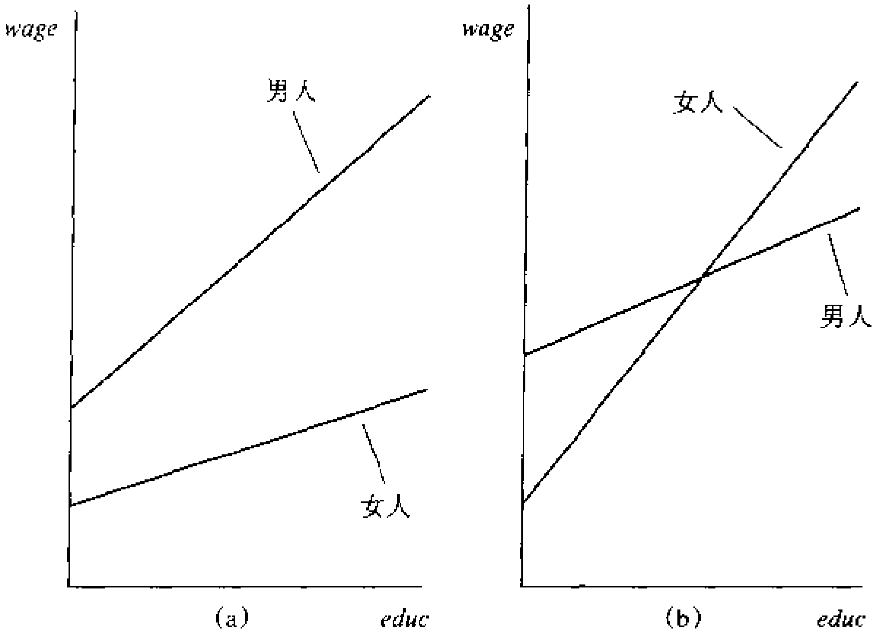
容许出现不同的斜率

226 我们现在已经看到了几个例子，表明在多元回归模型中容许任意几个组之间出现不同的斜率。在有些情况下，虚拟变量也可能与那些非虚拟的解释变量有交互作用，使得出现**不同的斜率**（differences in slopes）。继续看工资一例。假设在男人和女人的工资之间存在着恒定的差别的情况下（我们已经得到这种差别的证据），我们还想检验男人和女人受教育的回报是否相同。为简单起见，在模型中只包括受教育水平和性别。哪种模型会既存在恒定的工资差别又存在受教育回报上的差别呢？考虑模型

$$\log(wage) = (\beta_0 + \delta_0 female) + (\beta_1 + \delta_1 female)educ + u \qquad (7.16)$$

如果在模型（7.16）中代入 $female=0$，就会发现，男人这一组的截距是 β_0，而受教育的斜率是 β_1。对于女性，则代入 $female=1$；于是其截距是 $\beta_0+\delta_0$，而斜率是 $\beta_1+\delta_1$。所以，δ_0 度量了男人和女人在截距上的差异，而 δ_1 度量了男人和女人在受教育回报上的差异。δ_0 和 δ_1 的符号有四种情形，图7.2给出了两种

227 图（a）表明了女人组的截距小于男人组，而且女人组直线的斜率也小于男人组的情形。这意味着各种受教育程度的女人挣得都比男人少，而且其工资差距随着 *educ* 的提高而扩大。图(b)表明了女人组的截距小于男人组，



(a) $\delta_0 < 0,\ \delta_1 < 0$ (b) $\delta_0 > 0,\ \delta_1 > 0$

图 7.2 方程（7.16）的图示

但女人组直线的斜率却大于男人组的情形。这意味着女人在受教育水平很低的时候挣得比男人少，但随着受教育水平的提高，工资差距会逐渐缩小。到了一定的程度后，在给定相同的受教育水平的情况下，女人挣得可能比男人多（给定估计方程，这一点很容易求得。）

如何估计模型（7.16）呢？为了应用 OLS，我们必须写成一个含有 *female* 和 *educ* 乘积项的模型：

$$\log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female \cdot educ + u \qquad (7.17)$$

现在，就可以从$\log(wage)$对 *female*，*educ* 和 *female·educ* 的回归中估计出参数。在任何一个回归软件包中都能很容易地得到这个乘积项。不要被 *female·educ* 奇怪的性质所吓倒，对于样本中的每个男人，它都等于零，而对于样本中的每个女人，它都等于其受教育水平。

一个重要的假设是，男人和女人受教育的回报是相同的。就模型（7.17）而论，它被表述成 H_0：$\delta_1 = 0$，它意味着$\log(wage)$对 *educ* 的斜率在男人和女人之间是相同的。注意，这个假设对截距上的差异 δ_0 没有做任何限制。在这个虚拟假设之下，允许男人和女人之间存在工资差异，只是这种工资差异在各种相同的受教育水平上都必须相同。图 7.1 就描述了这种情况。

我们还对受教育水平相同的男人和女人的平均工资一样这个假设感兴趣。这意味着，在这个虚拟假设下，δ_0 和 δ_1 都必须同时为零。在方程（7.17）中，我们必须使用 F 检验来检验 H_0：$\delta_0 = 0$，$\delta_1 = 0$。而在只有截距差异的模型中，因为 H_1：$\delta_0 < 0$ 有力地拒绝了 H_0：$\delta_0 = 0$，所以我们拒绝过这个假设。

例 7.10　对数小时工资方程

我们在方程（7.17）中增加工作经历和现职任期的二次式：

$$
\begin{aligned}
\log(\widehat{wage}) = {} & 0.389 - 0.277\,female + 0.082\,educ - 0.005\,6\,female \cdot educ \\
& (0.119)(0.168) \qquad (0.008) \qquad (0.0131) \\
& + 0.029\,exper - 0.000\,58\,exper^2 + 0.032\,tenure \\
& (0.005) \qquad (0.000\,11) \qquad (0.007) \\
& - 0.000\,59\,tenure^2 \\
& (0.000\,24) \qquad\qquad\qquad\qquad (7.18) \\
& n = 526,\ R^2 = 0.441
\end{aligned}
$$

这个方程中估计男人的教育回报为 0.082 或 8.2%。女人的教育回报是 0.082 − 0.005 6 = 0.076 4 或约 7.6%。女人低 0.56% 或刚超过半个百分点，在经济上不大，在统计上也不显著：t 统计量为 −0.005 6/0.013 1 ≈ −0.43。228 我们因此断定，没有证据能够拒绝男人和女人具有相同的教育回报这个假设。

female 的系数尽管在经济上仍然较大，但在通常的置信水平上不再显著（$t = -1.35$）。其系数及在不含交互项的方程中的 t 统计量分别是 −0.297 和 −8.25［参见方程（7.9）］。我们现在应该断定没有统计显著的证据拒绝女人在相同的 *educ*，*exper* 和 *tenure* 时得到较低的回报吗？这将是一个严重的错误。由于我们已经在方程中增加了交互项 *female*·*educ*，所以 *female* 的估计系数与在方程（7.9）中的估计系数相比要欠准确得多：标准误几乎提高了 5 倍（0.168/0.036 ≈ 4.67）。其原因在于，*female* 和 *female*·*educ* 在样本中高度相关。在此例中，有一种考虑多重共线性的可取方法：在方程（7.17）和方程（7.18）中所估计的更一般的方程中，δ_0 度量了男人和女人在 $educ = 0$ 时的工资差异。由于样本中没有一个人具有甚至是接近于零年的受教育水平，所以我们在估计 $educ = 0$ 时的工资差异的过程中一度出现困难也就不足为奇（受教育水平为零年时的工资差异也不是很有信息含量）。更有意义的做法是，比方说在样本的平均受教育水平（12.5）上估计性别差异。为此，我们将以 *female*·(*educ* − 12.5) 取代 *female*·*educ* 并重新进行回归；这只会改变 *female* 的系数及其标准误。（见习题 7.15。）

如果我们对 H_0：$\delta_0 = 0$，$\delta_1 = 0$ 计算 F 统计量，那我们就得到 $F = 34.33$，对于一个分子 $df = 2$ 和分母 $df = 518$ 的 F 随机变量而言，这是一个相当巨大的值：p 值在小数点后四位都是零。最后，我们宁可选择表明了男女间存在一个不变的工资差异的模型（7.9）。

问题 7.4 ☞

你如何扩充式（7.18）中所估计的模型，以便考虑现职任期的回报中能在性别上有所差异？

作为一个涉及交互项的更复杂的模型，我们现在来看看种族和城市的种族构成

对大型棒球俱乐部中的运动员薪水的影响。

例 7.11　种族对棒球运动员薪水的影响

针对 330 个大型俱乐部的棒球运动员估计了如下方程，并有其所在城市种族构成的统计量可供使用。变量 *black* 和 *hispan* 都是每个运动员的二值指标。（基组是白人运动员。）变量 *percblck* 是该队所处城市中黑人的百分比，而 *perchisp* 是西班牙裔的比例。其他变量则度量了运动员在生产力和资历的某些方面。这里，我们感兴趣的是在控制了这些其他变量后种族的影响。

除了在方程中包括 *black* 和 *hispan* 外，还增加了交互项和 *black*·*percblck* 和 *hispan*·*perchisp*。所估计的方程是

229

$$
\begin{aligned}
\log(\widehat{salary}) = {} & 10.34 + 0.067\,3 years + 0.008\,9 gamesyr \\
& (2.18)\quad (0.012\,9)\qquad (0.003\,4) \\
& + 0.000\,95 bavg + 0.014\,6 hrunsyr + 0.004\,5 rbisyr \\
& (0.001\,51)\qquad (0.016\,4)\qquad (0.007\,6) \\
& + 0.007\,2 runsyr + 0.001\,1 fldperc + 0.007\,5 allstar \\
& (0.004\,6)\qquad (0.002\,1)\qquad (0.002\,9) \\
& - 0.198 black - 0.190 hispan \qquad (7.19) \\
& (0.125)\qquad (0.153) \\
& + 0.012\,5 black \cdot percblck + 0.020\,1 hispan \cdot perchisp \\
& (0.005\,0)\qquad (0.009\,8) \\
& n = 330,\ R^2 = 0.638
\end{aligned}
$$

我们首先应该检验 *black*，*hispan*，*black*·*percblck* 和 *hispan*·*perchisp* 这四个种族变量是否联合显著。使用这同样的 330 个运动员，在去掉这四个种族变量后的 R^2 为 0.626。由于这里有四个约束，而且不受约束模型的 $df = 330 - 13$，所以 F 统计量约为 2.63，这就得到一个等于 0.034 的 p 值。所以，这些变量在 5% 的水平上是联合显著的（尽管在 1% 的水平上不是）。

我们如何解释这些种族变量的系数呢？在以下的讨论中，所有的生产力因素都保持不变。首先，在保持 *perchisp* 不变的情况下，看看黑人运动员会怎么样。*black* 的系数 −0.198 确实意味着，如果一名黑人运动员在一个没有黑人的城市里（$percblck = 0$），那么这个黑人比一个条件相当的白人少挣约 19.8%。随着 *percblck* 的提高（由于 *perchisp* 保持不变，所以这意味着白人减少），黑人的薪水相对白人的薪水逐渐增加。在一个拥有 10% 的黑人的城市里，黑人的 log(*salary*) 比白人小 $-0.198 + 0.012\,5(10) = -0.073$，所以在这样一个城市里，黑人的薪水比白人约少 7.3%。当 $percblck = 20\%$ 时，黑人挣得比白人还要高出 5.2%。黑人比例最高的城市达到 74%（底特律）。

类似地，西班牙人在那些西班牙人比例很低的城市里挣得也比白人少。但我们很容易就能得到使白人和西班牙人的工资差异为零的 *perchisp*：必须使 $-0.190 + 0.020\,1 perchisp = 0$，即 $perchisp = 9.45$。对于那些西班牙人比

例低于 9.45%的城市而言，预计西班牙人挣得比白人少（给定黑人人口数），反之，如果西班牙人数超过 9.45%，则预计西班牙人挣得比白人多。22 个样本城市中有 12 个城市的西班牙人占总人口的比例不足 6%。西班牙人的最大比例约为 31%。

如何解释这些结论呢？我们不能简单地宣称存在对黑人和西班牙人的歧视，因为在那些少数民族聚居的城市里，白人挣得比黑人和西班牙人还要少。城市种族构成对薪水的重要性可能源于运动员的偏好：可能最好的黑人运动员不成比例地居住在那些黑人较多的城市，而最好的西班牙籍运动员则倾向于居住在那些西班牙人较多的城市里。式（7.19）中的估计值使我们能够确定存在某种关系，但我们不能辨别这两个假设。

检验不同组之间回归函数上的差别

230 上例说明，虚拟变量与其他自变量的交互可成为一个强有力的工具。有时候，我们想检验的虚拟假设是，两个总体或两组具有同一个回归函数，而对立假设是，各组间有一个或多个斜率是不同的。我们在第 13 章讨论跨时横截面的混合时，还将看到这样的例子。

假设我们想检验是否有一个相同的回归模型来描述大学男女运动员的大学 *GPA*。这个方程是

$$cumgpa = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + u$$

式中，*sat* 为 *SAT* 分数；*hsperc* 为高中的排名百分位；*tothrs* 为大学课程的总学时数。我们知道，为了出现不同的截距，可以包括男性或女性的一个虚拟变量。如果我们想让某个斜率取决于性别，只须在方程中包括一个适当变量与（比方说）*female* 的乘积。

如果想检验男人和女人之间是否存在差异，就必须容许模型的截距和斜率在两组间都不同：

$$\begin{aligned} cumgpa = {} & \beta_0 + \delta_0 female + \beta_1 sat + \delta_1 female \cdot sat + \beta_2 hsperc \\ & + \delta_2 female \cdot hsperc + \beta_3 tothrs + \delta_3 female \cdot tothrs + u \end{aligned} \tag{7.20}$$

参数 δ_0 是女人组和男人组之间在截距上的差异，而 δ_1 则是男女之间在 *sat* 的斜率上的差异，等等。男人和女人的 *cumgpa* 都遵循同一个模型的虚拟假设表述为

$$H_0: \delta_0 = 0, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0 \tag{7.21}$$

如果这些 δ_j 中有一个异于零，这个模型在男人和女人之间就是不同的。

利用文件 GPA3.RAW 中春季学期的数据，完整的模型被估计为

$$
\begin{aligned}
cumgpa = {} & 1.48 - 0.353 female + 0.001\,1 sat + 0.000\,75 female \cdot sat \\
& (0.21)(0.411) \qquad (0.000\,2) \qquad (0.000\,39) \\
& - 0.008\,5 hsperc - 0.000\,55 female \cdot hsperc + 0.002\,3 tothrs \\
& (0.001\,4) \qquad (0.003\,16) \qquad (0.000\,9) \\
& - 0.000\,12 female \cdot tothrs \qquad (7.22) \\
& (0.001\,63) \\
& n = 366,\ R^2 = 0.406,\ \bar{R}^2 = 0.394
\end{aligned}
$$

女性虚拟变量很显著，但没有一个交互项是显著的；只有交互项 *female*·*sat* 的 t 统计量接近于 2。但我们知道，检验像式（7.21）这样的一个联合假设，最好不要使用个别 t 统计量。为了计算 F 统计量，我们必须估计去掉 *female* 和所有交互项后的受约束模型；这样就得到一个约等于 0.352 的 R^2（受约束的 R^2），因而 F 统计量约为 8.14；p 值在小数点后五位都等于零，这就使我们能有力地拒绝式（7.21）。因此，尽管式（7.22）中表明男女有别的每一项，个别地看，在 5%的显著性水平上都是不显著的，但男女运动员的 *GPA* 模型确实不同。 231

female 和交互项的标准误较大，这使我们很难准确地讲男女之间到底有什么不同。由于在得到男女差别时必须考虑交互项，所以我们必须极其小心地解释方程（7.22）。如果只看变量 *female*，我们将错误地得到如下结论：保持其他因素不变，女人的 *cumgpa* 将比男人少约 0.353。它只是在设 *sat*、*hsperc* 和 *tothrs* 都等于零时所得到的估计差异，而这种情况并不是很有意义。在 $sat = 1\,100$，$hsperc = 10$ 和 $tothrs = 50$ 时，预计女人和男人之间的差异为 $-0.353 + 0.000\,75(1\,100) - 0.000\,55(10) - 0.000\,12(50) \approx 0.461$。也就是说，预计女运动员的 *GPA* 比同等条件的男运动员约高出半分。

在一个包含 *sat*，*hsperc* 和 *tothrs* 三个变量的模型中，添加所有的交互项来检验组间差别则相当容易。在某些情况下，会涉及更多的变量，那么，另有一种计算统计量的方法也很方便。它表明，即使在涉及许多自变量时，也能很容易地计算 F 统计量的残差平方和。

在含有 k 个解释变量和一个截距项的一般模型中，假设有两组，称为 $g = 1$ 和 $g = 2$。我们想检验这两组的截距和所有的斜率都相同。对 $g = 1$ 和 $g = 2$，将模型写成

$$
y = \beta_{g,0} + \beta_{g,1}x_1 + \beta_{g,2}x_2 + \cdots + \beta_{g,k}x_k + u \qquad (7.23)
$$

假设式（7.23）中两组间的每个 β 都相同就产生了 $k+1$ 个约束（在 *GPA* 一例中，$k+1=4$）。我们可以认为不受约束模型除了截距和变量本身外，还有一组虚拟变量和交互项，那么其自由度就是 $n-2(k+1)$。［在 *GPA* 一例中，$n-2(k+1) = 366-2(4) = 358$。］迄今为止，还没有什么新东西。关键是要洞察到，不受约束模型的残差平方和可通过两个分离的回归得到，这两个不同回归分别对应着两个不同的组。令 SSR_1 表示针对第一组估计式（7.23）所得到的残差平方和，它涉及 n_1 个观测。令 SSR_2 表示针对第二组估计式（7.23）所得到的残差平方和（n_2 个观测）。在上例中，若第一组为

女性，则 $n_1 = 90$，$n_2 = 276$。现在，不受约束模型的残差平方和无非就是 $SSR_{ur} = SSR_1 + SSR_2$。而受约束模型的残差平方和也就是将两组混合并估计一个方程时所得到的 SSR。一旦我们得到了这些，就可以像平常那样计算 F 统计量：

$$F = \frac{SSR - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \cdot \frac{n - 2(k+1)}{k+1} \tag{7.24}$$

式中，n 为总观测次数。在计量经济学中，通常将这个特定的 F 统计量称为**邹至庄统计量**（Chow statistic）。

为了在 GPA 一例中应用邹至庄统计量，我们需要将两组混合之后做回归所得到的 SSR，即 $SSR_r = 85.515$。样本中 90 个女人的 SSR 为 $SSR_1 = 19.603$，而男人组的 SSR 则为 $SSR_2 = 58.752$。因此，$SSR_{ur} = 19.603 + 58.752 = 78.355$。$F$ 统计量就是 $[(85.515 - 78.355)/78.355](358/4) \approx 8.18$；当然，考虑到四舍五入的误差，这就是我们在包含和不包含交互项
232 的两个模型中，用 R-平方的检验形式所得到的数值结果。（提醒一句：如果对每一组都分别估计一个回归，就不存在简单的 R-平方的检验形式；只有通过包括交互项来构造不受约束模型时，才能使用 R-平方的检验形式。）

无论用什么方法进行邹至庄检验，它都有一个重要的局限，即虚拟假设要求各组之间不存在任何差异。在更多的情况下，容许组间的截距不同再来检验斜率的差别会更有意义；我们已经在例 7.10 的工资方程中看到了一个这样的例子。为此，我们必须将交互项直接放到方程中，并检验所有交互项的联合显著性（不去约束截距项）。在 GPA 一例中，取虚拟假设为 $H_0: \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$。（虚拟假设不对 δ_0 进行约束。）这三个约束的 F 统计量约为 1.53，它给出的 p 值等于 0.205。因此，我们不能拒绝虚拟假设。

不能拒绝交互项的参数都为零的假设，表明最好的模型就是只容许截距的不同：

$$\begin{aligned}\widehat{cumgpa} = {} & \underset{(0.18)}{1.39} + \underset{(0.059)}{0.310}\,female + \underset{(0.0002)}{0.0012}\,sat - \underset{(0.0012)}{0.0084}\,hsperc \\ & + \underset{(0.0007)}{0.0025}\,tothrs \\ & n = 366, R^2 = 0.398, \bar{R}^2 = 0.392\end{aligned} \tag{7.25}$$

式(7.25)中的斜率系数接近于式(7.22)中基组(男性)的斜率系数，去掉交互项几乎没有什么变化。但式(7.25)中的 *female* 是高度显著的：其 t 统计量超过了 5，而这个估计值意味着，在给定 *sat*，*hsperc* 和 *tothrs* 的水平时，预计一名女运动员的 GPA 要比一名男运动员的 GPA 高 0.31 分。这实际上是一个十分重要的差异。

7.5 二值因变量：线性概率模型

到目前为止，我们已经学习了多元线性回归模型的许多性质和应用。我们在以上几节中学习了如何通过二值自变量的使用，使得定性信息成为一个多元回归模型中的解释变量。在迄今为止的所有模型中，因变量 y 都具有定量的含义（比如，y 表示美元的数量、一项考试的分数、一种百分比，或这些变量的对数）。如果想用多元回归来解释一个定性事件，结果会怎么样呢？

233 在实践中经常遇到的最简单情形中，我们想解释的事件是二值结果。换句话说，因变量 y 只取 0 和 1 两个值。比如，可以定义 y 表示一个成年人是否受过高中教育；或者用 y 表示一个大学生在某给定的学年中是否用过非法的毒品；或者用 y 表示一个企业在某给定年份是否接管了另一个企业。在上述每一例中，都可以令 $y=1$ 表示一种结果，而 $y=0$ 表示另一种结果。

当我们写出

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k+u \tag{7.26}$$

这个多元回归模型（其中 y 是一个二值变量）时，它有什么含义呢？由于 y 只能取两个值，所以就不能把 β_j 理解为，在保持所有其他因素不变的情况下，给定 x_j 一个单位的提高，导致 y 的变化量：y 要么从 0 变化到 1，要么从 1 变化到 0。尽管如此，我们仍然能对 β_j 作出有用的解释。如果我们假定零条件均值假定 MLR.3 成立，即 $\mathrm{E}(u\mid x_1,\cdots,x_k)=0$，那么会像往常一样得到

$$\mathrm{E}(y\mid \mathbf{x})=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k$$

式中，$\mathbf{x}$ 为所有解释变量的简单记法。

关键的一点是，当 y 是一个取值 0 和 1 的二值变量时，“成功”的概率 $\mathrm{P}(y=1\mid \mathbf{x})=\mathrm{E}(y\mid \mathbf{x})$（即 $y=1$ 的概率）等于 y 的期望值总是成立的。于是，我们得到一个重要的方程

$$\mathrm{P}(y=1\mid \mathbf{x})=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k \tag{7.27}$$

它说明成功的概率 $p(\mathbf{x})=\mathrm{P}(y=1\mid \mathbf{x})$ 是 x_j 的一个线性函数。方程（7.27）是二值响应模型的一个例子，而 $\mathrm{P}(y=1\mid \mathbf{x})$ 也被称为**响应概率**（response probability）。（我们在第 17 章还将讨论其他的二值响应模型。）由于概率和必须等于 1，所以 $\mathrm{P}(y=0\mid \mathbf{x})=1-\mathrm{P}(y=1\mid \mathbf{x})$ 也是 x_j 的一个线性函数。

因为这个响应概率是参数 β_j 的线性函数，所以这种带有二值因变量的多元线性回归模型又被称为**线性概率模型**（linear probability model，LPM）。在 LPM 中，在保持其他因素不变的情况下，β_j 度量了因 x_j 的变化导致成功概率的变化：

$$\Delta P(y=1|x)=\beta_j\Delta x_j \tag{7.28}$$

有了这些，我们就能使用多元回归模型来估计各个解释变量对定性信息的影响。OLS 机制也和从前一样。

如果把所估计的方程写成

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2+\cdots+\hat{\beta}_kx_k$$

那我们现在必须记住，$\hat{y}$ 就是预计的成功概率。因此，$\hat{\beta}_0$ 就是在每个 x_j 都
234 等于零时预计的成功概率，它可能有意义，也可能没有什么意义。斜率系数 $\hat{\beta}_1$ 度量的是，当 x_1 提高一个单位时，成功概率的预期变化。

为了正确地解释线性概率模型，我们必须了解“成功”是由什么构成的。因此，最好给因变量取一个能描述事件 $y=1$ 的名字。举例而言，令 *inlf*（“参与劳动市场”）为表示已婚妇女在 1975 年的劳动力参与状况的一个二值变量：如果一位妇女报告称她在该年度的某个时候曾为了工资而在家庭以外工作过，则 $inlf=1$；否则，$inlf=0$。我们假定劳动力参与还取决于收入的其他来源，包括丈夫的收入（*nwifeinc*，以千美元计）、受教育年数（*educ*）、过去在劳动力市场的年数（*exper*）、年龄（*age*）、年龄低于 6 岁的子女数（*kidslt*6）和年龄介于 6 岁～18 岁的子女数（*kidsge*6）。利用姆罗兹（Mroz，1987）的数据，我们估计了如下线性概率模型，其中，753 个妇女的样本中有 428 个人曾在 1975 年的某个时间参加过劳动：

$$\begin{aligned}\widehat{inlf}=&\underset{(0.154)}{0.586}\ \ \underset{(0.0014)}{0.0034}\,nwifeinc+\underset{(0.007)}{0.038}\,educ+\underset{(0.006)}{0.039}\,exper\\&-\underset{(0.00018)}{0.00060}\,exper^2-\underset{(0.002)}{0.016}\,age-\underset{(0.034)}{0.262}\,kidslt6+\underset{(0.0132)}{0.0130}\,kidsge6\end{aligned} \tag{7.29}$$

$$n=753,\ R^2=0.264$$

利用通常的 t 统计量，式（7.29）中除 *kidsge*6 外的所有变量都是统计显著的，而且所有统计显著的变量都具有基于经济理论（或常识）所预期的影响。

为了解释这些估计值，我们必须记得，因变量的变化改变了 $inlf=1$ 的概率。比如，*educ* 的系数意味着，保持式（7.29）中所有的其他因素都不变，多受一年教育使参与劳动市场的概率提高 0.038。如果我们直接看这个方程，多受 10 年教育会使参加劳动力的概率提高 0.038（10）＝0.38，这在概率上是一个相当大的提高。图 7.3 描绘了劳动力参与概率与 *educ* 之间的关系。为便于说明，其他自变量固定在 $nwifeinc=50$，$exper=5$，$age=30$，$kidslt6=1$ 和 $kidsge6=0$ 的水平上。直到受教育年数达到 3.84 年，才会使预期的概率为负。由于样本中没有哪个妇女的受教育年数低于 5 年，所以这也不会引起太多的担心。所报告的受教育年数最高达到 17 年，这就使预期的概率达到 0.5。如果我们让其他自变量取不同的数值，预期概率的范围也会随之变化。但多受一年教育对劳动力参与概率的边际影响总是 0.038。

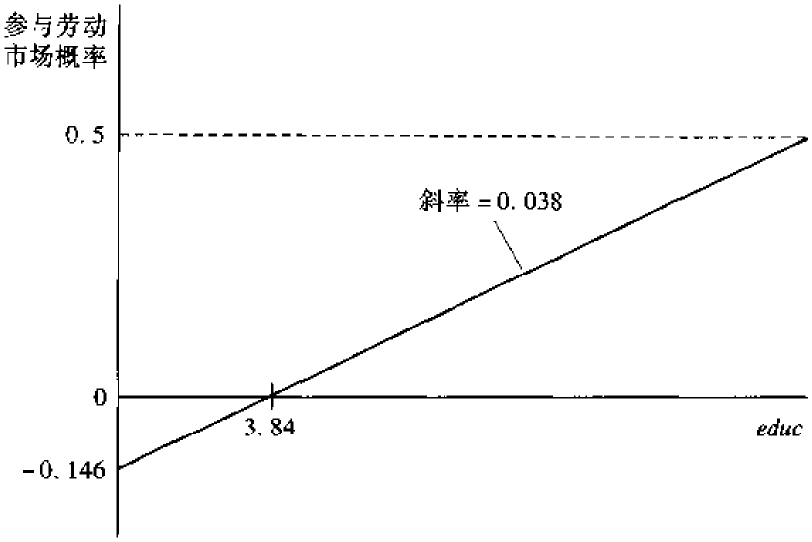


图 7.3　在其他解释变量不变的情况下，参与劳动市场的概率与受教育年数之间的估计关系

nwifeinc 的系数意味着，如果 $\Delta nwifeinc = 10$（意味着提高 10 000 美元），那么这位妇女参与劳动的概率就下降 0.034。由于以 1975 年美元计收入提高 10 000 美元是相当大的，所以这个影响不是特别大。工作经历以二次式出现，使得过去的工作经历对劳动力参与概率具有递减的影响。保持其他因素不变，概率的估计变化近似为 $0.039 - 2(0.000\,6)exper = 0.039 - 0.001\,2exper$。过去的工作经历对劳动力参与概率没有影响的点为 $0.039/0.001\,2 = 32.5$，这种工作经历的水平是很长的：样本中的 753 个妇女中，只有 13 个人的工作经历超过 32 年。

235 与较年长子女的个数不同，年幼子女的个数对劳动力参与具有巨大的影响。在给定其他变量水平的情况下，多一个不足 6 岁的子女，使参与劳动的概率减少 0.262。样本中，只有不足 20% 的妇女有一个或一个以上年幼子女。

此例说明了如何轻而易举地估计和解释线性概率模型，但它也表现出 LPM 的某些缺点。首先，很容易看到，如果在式（7.29）中代入自变量的某些特定组合数值，就能得到小于 0 或大于 1 的预测值。由于这些预测值都是概率，而概率必须介于 0～1 之间，所以这就有些尴尬。比如，预计一个妇女参与劳动的概率为 -0.10，它的含义是什么？实际上，样本中的 753 个妇女中，从式（7.29）中得到 16 个拟合值小于 0，17 个拟合值大于 1。

一个相关的问题是，概率不可能与自变量所有的可能值线性地相关。比如，式（7.29）预测，从 0 个子女增加到一个年幼子女的影响，使母亲参与劳动的概率下降 0.262。如果这位妇女从一个年幼子女增加到两个，那么概率的预期下降也是这么多。看起来更现实的情况是，第一个小孩使参与工作的概率下降很多，而以后增加的子女则具有越来越小的边际影响。实际上，极端地看，式（7.29）意味着，从 0 个年幼孩子增加到 4 个，使工作参与的概率减少 $\Delta inlf = 0.262(\Delta kidslt6) = 0.262(4) = 1.048$，但这是不可能的。

236 即使有这些问题，线性概率模型仍很有用处。线性概率模型常常应用于经济学，对自变量取值在样本均值附近特别奏效。在劳动力参与一例中，样本中没有一个妇女有 4 个孩子；实际上，只有三个妇女有 3 个孩子。96% 以上的妇女要么没有孩子，要么只有 1 个孩子。所以我们在解释所估计的方程时，也许应该仅关注这种情况。

在我们想做预测时，预测的概率超出单位区间会有一些问题，但这种情况很少会成为分析的核心。通常，我们想知道特定变量在其他条件不变的情况下对概率的影响。

由于 y 的二值特性，线性概率模型确实违背了一个高斯-马尔科夫假定。当 y 是一个二值变量时，其以 $\mathbf{x}$ 为条件的方差为

$$\mathrm{Var}(y \mid \mathbf{x}) = p(\mathbf{x})[1 - p(\mathbf{x})] \tag{7.30}$$

式中，$p(\mathbf{x})$ 为成功概率的简记：$p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$。这意味着，除非在概率与任何一个自变量都不相关，否则，线性概率模型中就一定存在着异方差性。我们从第 3 章知道，这不会导致 β_j 的 OLS 估计量出现偏误。但我们又从第 4 章和第 5 章知道，即使在大样本的情况下，同方差性对通常的 t 和 F 统计量是否正确而言都至关重要。我们还将在第 8 章说明，如何针对这种异方差性来修正标准误。最终表明，在许多应用研究中，通常的 OLS 统计量并非离谱，并且在应用研究中对线性概率模型进行标准的 OLS 分析仍是可接受的。

例 7.12　拘捕率的一个线性概率模型

令 $arr86$ 为一个二值变量，若一个人在 1986 年间曾被拘捕过则取值 1，否则取值 0。总体是 1960 年或 1961 年在加利福尼亚出生并在 1986 年以前至少被拘捕过一次的青年人群。刻画 $arr86$ 的一个线性概率模型就是

$$\begin{aligned} arr86 = {} & \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 tottime + \beta_4 ptime86 \\ & + \beta_5 qemp86 + u \end{aligned}$$

式中，$pcnv$ 为以前被捕后定罪的比例；$avgsen$ 为以前定罪后关进监狱的平均时间长度（以月为单位）；$tottime$ 为从 18 岁到 1986 年坐牢的总月数；$ptime86$ 为 1986 年坐牢的总月数；$qemp86$ 为此人 1986 年合法就业的季度数。

我们利用了 CRIME1.RAW 中的数据，这些数据在例 3.5 中也曾用过。现在，因为样本中只有 7.2% 曾不止一次被拘捕，我们使用了一个二值因变量。约 27.7% 的人在 1986 年被拘捕至少一次。所估计的方程是

237

$$\begin{aligned} \widehat{arr86} = {} & 0.441 - 0.162 pcnv + 0.006\ 1 avgsen \\ & (0.017)(0.021) \qquad (0.006\ 5) \\ & - 0.002\ 3 tottime - 0.022 ptime86 - 0.043 qemp86 \\ & (0.005\ 0) \qquad (0.005) \qquad (0.005) \\ & n = 2\ 725, R^2 = 0.047\ 4 \end{aligned} \tag{7.31}$$

截距 0.441 表示了一个从未定过罪（所以 *pcnv* 和 *avgsen* 都等于零）、18 岁以后从未坐过牢、1986 年也没有进过监狱而且整个 1986 年从未就业的人预计会被拘捕的概率。无论个别地看还是联合地看，变量 *avgsen* 和 *tottime* 都不显著（F 统计量给出的 p 值为 0.347），而且如果说更长期的判刑能阻止犯罪的话，*avgsen* 的符号则有些违背我们的直觉。格罗戈尔（Grogger, 1991）利用这些数据的一个超集和一个不同的计量方法，发现 *tottime* 对拘捕概率有统计显著的正影响，并认为 *tottime* 就是对犯罪活动中形成的人力资本的一种度量。

虽然定罪概率的提高确实能降低拘捕概率，但我们在解释这个系数的大小时必须小心。变量 *pcnv* 是介于 0～1 之间的一个比例，于是，*pcnv* 从 0 变化到 1 实际上意味着从没有可能被定罪到必然被定罪。即使是这么大的变化，也只能使拘捕概率减小 0.162；*pcnv* 提高 0.5 使拘捕概率减小 0.081。

监禁的影响由 *ptime86* 的系数给出。如果一个人在坐牢，那他就不可能被拘捕。由于 *ptime86* 是以月为单位度量的，所以在监狱里多呆 6 个月，会使拘捕概率减少0.022(6) = 0.132。方程（7.31）还给出了线性概率模型不是对所有自变量值都成立的另一个例子。如果一个人在 1986 年的 12 个月里都在坐牢，那他在 1986 年就不可能被拘捕。取所有其他变量的值为零，在 *ptime86* = 12 时预期被拘捕的概率为0.441 − 0.022(12) = 0.177，而不是零。不过，如果我们从拘捕的无条件概率 0.277 开始，那么，12 个月的监禁则使拘捕概率基本下降到零：0.277 − 0.022(12) = 0.013。

最后，就业也显著降低拘捕概率。在所有其他因素不变的情况下，一个四个季度都在工作的人，与一个完全不工作的人相比，被拘捕的可能性降低 0.172。

我们还可以在含有虚拟因变量的模型中引入虚拟自变量，其系数度量了虚拟变量从 0 变化到 1 而导致成功概率的预期变化。比如，如果我们在拘捕方程中增加两个种族虚拟变量 *black* 和 *hispan*，则得到

288

$$\begin{aligned}
\widehat{arr}86 = {} & 0.380 - 0.152pcnv + 0.0046avgsen - 0.0026tottime \\
& (0.019)\quad (0.021) \qquad (0.0064) \qquad (0.0049) \\
& - 0.024ptime86 - 0.038qemp86 + 0.170black \\
& (0.005) \qquad\qquad (0.005) \qquad\quad (0.024) \\
& + 0.096hispan \\
& (0.021) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (7.32) \\
& n = 2\,725,\ R^2 = 0.0682
\end{aligned}$$

问题 7.5 一个没有案底（所以 *pcnv*、*avgsen*，*tottime* 和 *ptime86* 都等于零）而且在 1986 年对四个季度都在就业的黑人，预计其被拘捕的概率是多大？这个结论合理吗？

式中的系数意味着，在所有其他因素保持不变的情况下，一个黑人比一个白人（基组）被拘捕的概率要高出 0.17。换个说法就是，黑人比白人被

拘捕的概率高 17 个百分点。这个差别还是统计显著的。类似地，西班牙人比白人被拘捕的概率也要高 0.096。

7.6 对政策分析和项目评价的进一步讨论

我们已经看到一些可用于政策评价的含虚拟变量的模型。例 7.3 给出了一个项目评价的例子，其中某些企业得到了在职培训津贴，而其他企业则没有。

像我们前面提到的那样，由于在社会科学的多数例子中，对照组和处理组并不是随机指定的，所以在评价一个项目时还必须十分小心。再次考虑霍尔泽等人（Holzer et al., 1993）的研究，我们现在感兴趣的是在职培训津贴对工人生产力（而不是在职培训的数量）的影响。我们所关心的方程是

$$\log(scrap) = \beta_0 + \beta_1 grant + \beta_2 \log(sales) + \beta_3 \log(employ) + u$$

式中，*scrap* 为企业的废弃率；后两个变量作为控制变量包括进来。二值变量 *grant* 表示企业在 1988 年是否得到了在职培训津贴。

在看估计值之前，我们可能担心、一些影响工人生产力的观测不到的因素（如所受教育、能力、工作经历和现职任期的平均水平等）可能会与企业能否得到津贴相关。霍尔泽等人指出，津贴采取先到先供应的方式发放。但这并非等同于随机发放。工人生产力较低的企业可能看到了提高生产力的机会，于是更加致力于申请津贴。

利用 JTRAIN.RAW 中 1988 年的数据——当时企业实际上更适合于得到津贴，我们得到

$$\begin{aligned}\log(\widehat{scrap}) = {} & \underset{(4.66)}{4.99} - \underset{(0.431)}{0.052}\,grant - \underset{(0.373)}{0.455}\log(sales) \\ & + \underset{(0.365)}{0.639}\log(employ) \qquad (7.33) \\ & n = 50, R^2 = 0.072\end{aligned}$$

(50 个企业中的 17 个得到了培训津贴，而所有企业的平均废弃率为 3.47%。) *grant* 的点估计值 −0.052 意味着，在给定 *sales* 和 *employ* 的情况下，得到津贴的企业的废弃率比没有得到津贴的企业约低 5.2%。在培训津贴有效的情况下，虽然估计的方向与预期影响的方向一致，但其 t 统计量相 239 当小。因此，通过对这个横截面的分析，我们必须承认，津贴对企业的生产力没有影响。我们在第 9 章还会回到这个例子上来，并说明如何通过增加上一年度的信息而导致极为不同的结论。

即便在政策分析没有涉及将各个单位指派到对照组和处理组的情形中，我们也必须警惕包含了那些可能与所关心的二值自变量系统相关的因素。这

方面的一个很好的例子就是对种族歧视的检验。种族是一个不能由个人和政府官员所决定的因素。实际上，一个人在出生时就决定了其种族，所以种族看上去是外生变量的一个很好的例子。然而，基于历史原因，事实并非如此：各种族之间在背景上存在着系统的差异，而这些差异在检验当前的歧视时又相当重要。

作为一个例子，考虑对贷款许可中歧视问题的检验。如果能搜集到个人抵押贷款申请方面的数据，那么我们就能定义一个虚拟变量 *approved*：若申请得到批准，则取值 1，否则取值 0。各种族在批准率上的系统差异就是歧视的一个指标。但由于批准贷款取决于许多其他因素，包括收入、财富、信用等级和偿还贷款的一般能力等，所以，如果这些因素在各种族间存在着系统差异，就必须对其加以控制。一个检验歧视问题的线性概率模型可能就具有如下形式：

$$approved = \beta_0 + \beta_1 nonwhite + \beta_2 income + \beta_3 wealth$$
$$+ \beta_4 credrate + \text{其他因素}$$

由于 β_1 表示的是给定方程中其他因素的水平不变，非白人得到许可的概率与白人得到许可的概率之间差异的大小，所以，拒绝 H_0：$\beta_1 = 0$ 而支持 H_1：$\beta_1 < 0$ 就表明对少数民族存在歧视。如果 *income*，*wealth* 等在各种族间存在系统差异的话，那么，在一个多元回归分析中对这些因素加以控制就很重要。

政策和项目评价中时常出现的另一个问题是，个人（或企业或城市）选择是否参与某种特定的行为和项目。比如，个人对使用非法毒品和喝酒的选择。如果我们想考查这种行为对失业状况、收入和犯罪行为的影响，我们就应该考虑，毒品的使用可能与其他能影响就业和犯罪结果的因素相关。适合于智力开发一类项目的孩子，其是否参与取决于父母的决策。由于家庭背景在智力开发决策中有重要作用并影响学生的未来，所以我们在考查智力开发项目的影响时，应该控制这些因素［参见 Currie and Thomas（1995）］。雇主和政府机构选拔参与在职培训项目的个人，都既可以参与又可以不参与，而是否参与不太像是随机的［参见 Lynch（1991）］。城市和州也可以选择是否实施某特定的枪支管制法律，而且这个决策也可能与那些影响暴力犯罪的因素系统相关［参见 Cleck and Patterson（1993）］。

240 上一段给出的几个例子在经济学中一般被称为**自选择**（self-selection）问题。照字面理解，这个词来自个人自己选择加入某种行为或项目的事实：参与并不是随机决定的。这个术语一般用于参与的二值指标可能与无法观测因素系统相关的情况。于是，如果我们写出一个简单的模型

$$y = \beta_0 + \beta_1 partic + u \tag{7.34}$$

式中，y 为一个结果变量；*partic* 为一个二值变量。如果一个人、一个企业或一个城市参与一种行为、一个项目或有某种法律，这个二值变量就取值 1。这样一来，我们担心的是，u 的平均值取决于参与决策：$\mathrm{E}(u \mid partic =$

$1)\neq \mathrm{E}(u \mid partic=0)$。正如我们所知，这会导致 β_1 的简单回归估计量产生偏误，所以我们不能发现参与的真正影响。因此，自选择问题是说明解释变量（这里就是 *partic*）能够内生的另一种方法。

到目前为止，我们知道多元回归分析在某种程度上可以缓解自选择问题。式（7.34）中误差项所包含的与 *partic* 相关的因素可以包括在一个多元回归方程中，当然，需要假定我们能搜集这些因素的数据。不幸的是，在许多情形中，我们担心观测不到因素与参与变量相关，在这种情况下，多元回归会导致有偏误的估计量。

在利用横截面数据进行多元回归分析时，我们必须警惕因自选择问题而得到项目对结果变量的谬误影响。柯里和科尔（Currie and Cole，1993）的研究中就包含了一个很好的例子。这两位作者考查了参与 AFDC（对有子女家庭的援助计划）对孩子出生时重量的影响。即便在控制了家庭和背景特征的一系列变量之后，作者得到的 OLS 估计量仍意味着参与 AFDC 会降低孩子出生体重。正像作者所指出的那样，很难相信 AFDC 参与本身会导致出生体重降低。[另外一个例子，可参见 Currie（1995）。] 利用我们在第 15 章将讨论的另一种计量经济方法，柯里和科尔就发现参与 AFDC 对出生体重没有影响或有正影响的证据。

当自选择问题因缺乏充分的控制变量而使多元回归分析出现偏误时，就可以使用第 13、14 章和第 15 章所讨论的更高级的方法。

▶ 小结

我们在本章了解到如何在回归分析中使用定性信息。在最简单的情形中，定义一个虚拟变量来区别两个组，而这个虚拟变量的系数估计了在其他条件不变情况下这两组之间的差异。若多于两个组，则可以定义一系列虚拟变量：若有 g 个组，则模型中就要包括 $g-1$ 个虚拟变量。对所有虚拟变量估计值的解释，都是相对于基组或基准组（模型中没有包含其虚拟变量的那一组）而言。

虚拟变量还可以用于在回归模型中包括诸如信用等级或选美排名等序数信息。我们只需要定义一系列虚拟变量来表示序数变量的不同结果，并允许其中的某个等级作为基组。

虚拟变量也能与定量变量相互作用，使不同组之间出现不同的斜率。在极端情形中，我们能够让每一组都既有自己的截距，又有自己的斜率。邹至庄检验可用来检查各组之间是否有差异。多数情况下，在允许两个不同的组之间存在截距上的差异的情况下，检验其斜率是否相同就更有意义。在
241 一个包括了虚拟变量组和所有变量之交互作用的不受约束模型中，使用标准的 F 统计量就能做这种检验。

仅用 OLS 估计的线性概率模型，使我们能够利用回归分析来解释二值

响应。现在，OLS估计值可理解为，给定对应解释变量变化一个单位，“成功”（$y=1$）的概率有多大的变化。LPM确实也有一些缺陷：它可能导致所预测的概率小于0或大于1；它意味着，每个以其原有形式出现的解释变量的边际效应为常数；而且它还包含了异方差性。如果我们要得到解释变量在数据中间部分的偏效应的估计值，那么前两个问题通常就不算严重。虽然异方差性确实使我们不能使用通常的OLS标准误和检验统计量，但如我们在下一章将见到的那样，在样本足够大的情况下，问题将迎刃而解。

本章在结束时讨论了如何用二值变量来评价政策和项目的问题。如在所有的回归分析中一样，我们必须记住，项目参与或其他某个具有政策含义的二值回归元，都可能会与那些观测不到而又影响因变量的因素相关，从而导致通常的变量遗漏偏误。

关键术语

基组
基准组
二值变量
邹至庄统计量
对照组
斜率差异
虚拟变量陷阱
虚拟变量
实验组
交互（作用）项
截距迁移
线性概率模型（LPM）
序数变量
政策分析
项目评价
响应概率
自选择
处理组

习 题

7.1 利用SLEEP75.RAW中的数据（也可参见习题3.3），我们得到如下估计方程：

$$\begin{aligned}\widehat{sleep} = {} & 3\,840.83 - 0.163\,totwrk - 11.71\,educ - 8.70\,age + 0.128\,age^2 \\ & (235.11)\quad (0.018)\qquad (5.86)\qquad (11.21)\quad (0.134) \\ & + 87.75\,male \\ & (34.3) \\ & n = 706,\ R^2 = 0.123,\ \bar{R}^2 = 0.117\end{aligned}$$

242 式中，*sleep* 为每周晚上睡眠的总分钟数；*totwrk* 为每周花在工作上的总分钟数；*educ* 和 *age* 则以年为单位；*male* 为一个性别虚拟变量。

(i) 所有其他因素不变，有没有男人比女人睡眠更多的证据？这个证据有多强？

(ii) 工作与睡眠之间有统计显著的取舍关系吗？所估计的取舍关系是什么样的？

(iii) 为了检验年龄在其他因素不变的情况下对睡眠没有影响这个虚拟假设，你还需要另外做什么回归？

7.2 利用 BWGHT.RAW 中的数据，可估计出如下方程：

$$
\begin{aligned}
\widehat{\log(bwght)} = {} & 4.66 - 0.0044\,cigs + 0.0093\log(faminc) \\
& (0.22)\ (0.0009) \qquad (0.0059) \\
& + 0.016\,parity + 0.027\,male + 0.055\,white \\
& (0.006) \qquad\quad (0.010) \qquad (0.013) \\
& n = 1388, R^2 = 0.0472
\end{aligned}
$$

和

$$
\begin{aligned}
\widehat{\log(bwght)} = {} & 4.65 - 0.0052\,cigs + 0.0110\log(faminc) \\
& (0.38)\ (0.0010) \qquad (0.0085) \\
& + 0.017\,parity + 0.034\,male + 0.045\,white \\
& (0.006) \qquad\quad (0.011) \qquad (0.015) \\
& + 0.0030\,motheduc + 0.0032\,fatheduc \\
& (0.0030) \qquad\qquad (0.0026) \\
& n = 1191, R^2 = 0.0493
\end{aligned}
$$

变量的定义和习题 4.9 中一样，但我们增加了两个虚拟变量：一个虚拟变量表明孩子是不是男孩，另一个虚拟变量则表明这个孩子是不是白人。

(i) 在第一个方程中，*cigs* 解释变量的系数。具体而言，每天多吸 10 根烟对出生体重有何影响？

(ii) 在第一个方程中，保持其他因素不变，预计一个白人孩子的出生体重比一个非白人孩子重多少？这个差异是统计显著的吗？

(iii) 评价 *motheduc* 的估计影响和统计显著性。

(iv) 从这些给定的信息中，为什么不能计算出检验 *motheduc* 和 *fatheduc* 联合显著性的 F 统计量？为了计算这个统计量，还需要做些什么？

7.3 利用 GPA2.RAW 中的数据，可估计出如下方程：

$$
\begin{aligned}
\widehat{sat} = {} & 1028.10 + 19.30\,hsize - 2.19\,hsize^2 - 45.09\,female \\
& (6.29) \qquad (3.83) \qquad (0.53) \qquad (4.29) \\
& - 169.81\,black + 62.31\,female \cdot black \\
& (12.71) \qquad\quad (18.15) \\
& n = 4137, R^2 = 0.0858
\end{aligned}
$$

243 式中，*sat* 为 *SAT* 的总分；*hsize* 为以百人计学生所在高中毕业年级的学生规模；*female* 为一个性别虚拟变量；*black* 为一个种族虚拟变量（黑人取值 1，其他人都取值 0）。

(i) 有很强的证据支持，模型中应该包括 $hsize^2$ 吗？从这个方程来看，最优的高中规模是什么？

(ii) 保持 $hsize$ 不变，非黑人女性和非黑人男性之间 SAT 分数的估计差异是多少？这个估计差异的统计显著性如何？

(iii) 非黑人男性和黑人男性之间 SAT 分数的估计差异是多少？检验其分数没有差异的虚拟假设，对立假设是他们的分数存在差异。

(iv) 黑人女性和非黑人女性之间 SAT 分数的估计差异是多少？为了检验这个差异的统计显著性，你需要怎么做？

7.4　一个解释 CEO 薪水的方程为

$$
\begin{aligned}
\log(\widehat{salary}) = {} & \underset{(0.30)}{4.59} + \underset{(0.032)}{0.257}\log(sales) + \underset{(0.004)}{0.011}\,roe \\
& + \underset{(0.089)}{0.158}\,finance + \underset{(0.085)}{0.181}\,consprod - \underset{(0.099)}{0.283}\,utility \\
& n = 209, R^2 = 0.357
\end{aligned}
$$

所用到的数据来自 CEOSAL1.RAW，其中 $finance$，$consprod$ 和 $utility$ 都是二值变量，分别表示金融业、消费品工业和公用事业。省略的产业部门是交通运输业。

(i) 保持 $sales$ 和 roe 不变，计算公用事业和交通运输业之间估计薪水的近似百分比差异。这个差异在 1% 的显著性水平上是统计显著的吗？

(ii) 利用方程 (7.10)，求出公用事业和交通运输业之间估计薪水的精确百分比差异，并与第 (i) 部分所得到的答案相比较。

(iii) 消费品工业和金融业之间估计薪水的近似百分比差异是多少？写出一个使你能检验这个差异是否统计显著的方程。

7.5　在例 7.2 中，令 $noPC$ 表示一个虚拟变量，没有一台个人计算机的学生取值 1，否则取值 0。

(i) 如果用 $noPC$ 取代方程 (7.6) 中的 PC，所估计方程中的截距会怎么样？$noPC$ 的系数是多少？［提示：写出 $PC = 1 - noPC$，并代入方程 $\widehat{colGPA} = \hat{\beta}_0 + \hat{\delta}_0 PC + \hat{\beta}_1 hsGPA + \hat{\beta}_2 ACT$。］

(ii) 如果用 $noPC$ 取代 PC，R-平方会有什么变化？

(iii) PC 和 $noPC$ 应该都作为自变量包括进模型中吗？请解释。

7.6　为了检验在职培训项目对工人以后工资的有效性，我们设定了模型

244

$$\log(wage) = \beta_0 + \beta_1 train + \beta_2 educ + \beta_3 exper + u$$

式中，$train$ 表示的虚拟变量在工人参与这个项目时取值 1。想像误差项包括了无法观测的工人能力。如果能力越低的工人被选派去参加这个项目的机会就越大，而且你使用了一个 OLS 分析，那你认为 β_1 的 OLS 估计量可能有什么样的偏误？［提示：参考第 3 章。］

7.7　在方程 (7.29) 的例子中，假设我们定义 $outlf$ 在妇女不属于劳

动力范围时等于 1，否则等于 0。

（i）如果我们将 *outlf* 对方程（7.29）中所有的自变量做回归，截距和斜率的估计值会怎么样？［提示：$inlf = 1 - outlf$。将它代入总体方程 $inlf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \cdots$，并重新整理。］

（ii）截距和斜率的标准误将有什么变化？

（iii）*R*-平方会有什么变化？

7.8　假设你通过对工资、受教育程度、工作经历和性别的调查来搜集数据，而且你还询问了大麻使用方面的信息。原问题是：“上个月你吸过几次大麻？”

（i）写出一个方程，使之在控制其他因素的情况下，能让你估计出使用大麻对工资的影响。你应该能得出这样的结论：“每个月多吸 5 次大麻，估计会改变工资 *x*%。”

（ii）写出一个模型，使你能检验女人和男人在使用大麻对工资的影响上是否存在差异。你将怎样检验男女之间在使用大麻的影响上是没有差异的？

（iii）假设你认为最好按大麻使用量将人分为四类：不用者、浅尝者（每月 1～5 次）、适度者（每月 6～10 次）和重用者（每月 10 次以上）。写出一个模型，使你能估计出使用大麻对工资的影响。

（iv）利用第（iii）部分的模型，详细解释如何检验使用大麻对工资没有影响的虚拟假设。既要具体，又要包括对自由度的一个仔细列表。

（v）利用你搜集来的调查数据做因果推断，会有哪些潜在的问题？

计算机习题

7.9　本题使用 GPA1.RAW 中的数据。

245 （i）在估计方程（7.6）中增加变量 *mothcoll* 和 *fathcoll*，并以通常的形式报告结果。拥有 PC 的估计影响会怎么样？PC 还是统计显著的吗？

（ii）检验第（i）部分方程中的联合显著性，不要忘记报告 *p* 值。

（iii）在第（i）部分的模型中增添 $hsGPA^2$，并决定是否进行这种增添。

7.10　本题使用 WAGE2.RAW 中的数据。

（i）估计模型

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + \beta_7 urban + u$$

并以通常的形式报告结果。保持其他因素不变，黑人和非黑人之间月薪的近似差异是多少？这个差异是统计显著的吗？

（ii）在这个方程中增加变量 $exper^2$ 和 $tenure^2$，证明即便在 20% 的显著性水平上，它们也是联合不显著的。

（iii）推广原模型，使受教育回报取决于种族，并检验受教育的回报是否的确取决于种族。

（iv）再回到原模型，但现在容许四个不同人群（已婚黑人、已婚非黑人、单身黑人和单身非黑人）的工资有差别。估计已婚黑人和已婚非黑人之

间的工资差异是多少？

7.11 一个容许大型棒球俱乐部运动员的薪水因位置不同而不同的模型是

$$\begin{aligned}\log(salary) = {} & \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr \\ & + \beta_5 rbisyr + \beta_6 runsyr + \beta_7 fldperc + \beta_8 allstar \\ & + \beta_9 frstbase + \beta_{10} scndbase + \beta_{11} thrdbase + \beta_{12} shrtstop \\ & + \beta_{13} catcher + u\end{aligned}$$

其中外场手为基组。

（i）表述如下虚拟假设：在控制了其他因素后，接球手和外场手的所得大致相同。利用 MLB1.RAW 中的数据检验这个假设，并评论所估计的薪水差异的大小。

（ii）表述并检验如下虚拟假设：一旦控制了其他因素，各个位置的平均薪水没有差别。

（iii）第（i）和第（ii）部分的结论相一致吗？如果不一致，解释为什么。

7.12 本题使用 GPA2.RAW 中的数据。

（i）考虑方程

$$\begin{aligned}colgpa = {} & \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female \\ & + \beta_6 athlete + u\end{aligned}$$

式中，*colgpa* 为累积的大学 *GPA*；*hsize* 为高中毕业年级以百人计的规模；*hsperc* 为在毕业年级中学术排名的百分位；*sat* 为总的 *SAT* 分数；*female* 为一个二值变量；*athlete* 也是一个运动员取值 1 的二值变量。你对这个方程中的系数有何预期？哪些你没有把握？

246 （ii）估计第（i）部分中的方程，并以通常的形式报告结果。估计运动员和非运动员之间 *GPA* 的差异是多少？它是统计显著的吗？

（iii）从模型中去掉 *sat* 并重新估计这个方程。现在，作为运动员的估计影响是多大？讨论为什么这个估计值不同于第（ii）部分的结论。

（iv）在第（i）部分的模型中，容许作为运动员的影响会因性别不同而不同。检验男女运动员在其他条件不变的情况下没有差别的虚拟假设。

（v）*sat* 对 *colgpa* 的影响会因性别不同而不同吗？讲出你的根据。

7.13 在习题 4.2 中，我们在一个解释 CEO 薪水的模型中增加了变量企业股票的回报 *ros*；结果表明，*ros* 是不显著的。现在，定义一个虚拟变量 *rosneg*，它在 $ros<0$ 时等于 1，而在 $ros\geqslant 0$ 时等于 0。利用 CEOSAL1.RAW 来估计模型

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 rosneg + u$$

讨论对 $\hat{\beta}_3$ 的解释及其统计显著性。

7.14 本题利用 SLEEP75.RAW 中的数据。所考虑的方程是

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + u$$

（i）分别对男人和女人估计这个方程，并以通常的形式报告结果。这两个估计方程有显著差异吗？

（ii）计算上述两个方程中参数相等的邹至庄检验。使用添加了 $male$ 和交互项 $male \cdot totwrk$，…，$male \cdot yngkid$ 的检验形式，并利用全部观测集。这个检验适当的 df 是多少？你会在 5% 的显著性水平上拒绝这个虚拟假设吗？

（iii）现在，容许男女之间存在着不同的截距，判断涉及 $male$ 的交互项是否联合显著。

（iv）给定第（ii）部分和第（iii）部分中得到的结论，你最后的模型是什么？

7.15 本题使用 WAGE1.RAW 中的数据。

（i）利用方程（7.18）估计在 $educ = 12.5$ 时的性别差异，并与 $educ = 0$ 时的性别差异相比较。

（ii）做一个用以得到方程（7.18）的回归，但用 $female \cdot (educ - 12.5)$ 取代 $female \cdot educ$。你现在如何解释 $female$ 的系数？

（iii）第（ii）部分中 $female$ 的系数是统计显著的吗？与方程（7.18）相比较并进行评论。

7.16 本题使用 LOANAPP.RAW 中的数据。要解释的二值变量是 $approve$，如果一个人的抵押贷款得到许可则取值 1。主要的解释变量是一个虚拟变量 $white$，如果申请者是白人则取值 1。数据集中其他的申请者为黑人和西班牙人。

247 为了检验抵押贷款市场中的歧视，可使用一个线性概率模型

$$approve = \beta_0 + \beta_1 white + \text{其他因素}$$

（i）如果对少数民族存在歧视并控制了适当的因素，那么，β_1 的符号是什么？

（ii）将 $approve$ 对 $white$ 做回归，并以通常的形式报告结果。解释 $white$ 的系数。它是统计显著的吗？它实际上大吗？

（iii）作为控制因素，增加变量 $hrat$，$obrat$，$loanprc$，$unem$，$male$，$married$，dep，sch，$cosign$，$chist$，$pubrec$，$mortlat1$，$mortlat2$ 和 vr。$white$ 的系数会有什么变化？仍有对非白人存在歧视的证据吗？

（iv）现在考虑种族与度量了其他债务占收入之比例的变量（$obrat$）存在着交互作用。交互项显著吗？

（v）利用第（iv）部分的模型，在债务负担达到样本均值 $obrat = 32$ 时，作为白人对贷款许可的概率有多大的影响？构造这种影响的一个 95% 的置信区间。

第 8 章　异方差性

248　第 3 章多元回归分析中引入的同方差性假定表明，以解释变量为条件（观测不到的）的误差方差是常数。只要不可观测因素的方差随总体的不同部分而变化（由不同的解释变量值所决定），同方差性就不能成立。比如，在一个储蓄方程中，如果影响储蓄而又观测不到的因素的方差随收入而变化，那就会出现异方差性。

我们在第 3 章和第 4 章看到，即便样本容量很大，要在线性回归模型中使用OLS估计的 t 检验、F 检验和置信区间，都需要同方差性假定。我们在本章讨论出现异方差性时的一些补救措施，并说明如何检验异方差性的出现。我们先从简要评论异方差性对普通最小二乘估计所造成的后果开始。

8.1　异方差性对 OLS 所造成的影响

再次考虑多元线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{8.1}$$

在第 3 章，我们在前四个高斯－马尔科夫假定 MLR.1～MLR.4 下证明了

OLS估计量 $\hat{\beta}_0$，$\hat{\beta}_1$，$\hat{\beta}_2$，…，$\hat{\beta}_k$ 的无偏性。我们在第5章证明了，同样的这前四个假定意味着OLS的一致性。用误差方差表示为Var$(u|x_1,x_2,\cdots,x_k)=\sigma^2$的同方差性假定MLR.5，在证明OLS的无偏性和一致性的过程中并没有起到什么作用。重要的是记住，异方差性并不会导致β_j的OLS估计量出现偏误或产生不一致性，但诸如省略一个重要变量之类的情况出现则具有这种影响。

如果异方差性不会导致偏误和不一致性，那我们为什么还要引入它作为一个高斯-马尔科夫假定呢？回想第3章，估计量的**方差**Var$(\hat{\beta}_j)$在没有同方差性假定的情况下是有偏的。由于OLS标准误直接以这些方差为基础，所以它们都不能用来构造置信区间和t统计量。在出现异方差性时，通常普通最小二乘法的t统计量就不具有t分布，使用大样本容量也不能解决这个问题。类似地，F统计量也不再是F分布，而LM统计量也不服从一个渐近的χ平方分布。总之，在出现异方差性的情况下，我们在高斯-马尔科夫假定下用来检验假设的统计量都不再成立。 249

我们还知道，表明OLS是最优线性无偏估计的高斯-马尔科夫定理，关键是依靠同方差性的假定。如果Var$(u|\mathbf{x})$不是常数，OLS就不再是BLUE。此外，定理5.3中描述的一类估计量中，OLS也不再是渐近有效的。如我们8.4节中将看到的那样，在出现异方差性的情况下，可能会找到比OLS更有效的估计量（尽管这要求我们知道异方差性的形式）。在样本容量相对较大时，得到一个有效估计量可能就不是那么重要。我们在下一节将说明，如何修正通常的OLS检验统计量，并使之至少渐近有效。

8.2 OLS估计后异方差—稳健性推断

由于假设检验在计量经济分析中如此重要，而通常的OLS推断在出现异方差时一般都是错的，所以必须决定我们是否应该完全放弃OLS。幸运的是，OLS仍然有用。在最近20年间，计量经济学家已经知道了该如何调整标准误、t统计量、F统计量和LM统计量，使之在出现**未知形式的异方差性**（heteroskedasticity of unknown form）时仍有效。这是很方便的，因为它意味着无论总体出现的异方差性的类型如何，都能报告能起作用的新统计量。因为无论误差方差是否为常数，而且我们还不需要知道到底是哪种情况，它们都（至少在大样本下）是有效的，所以我们把本节讨论的这种方法称为**对异方差—稳健性程序**。

我们首先概括一下在出现异方差时如何估计方差Var$(\hat{\beta}_j)$。虽然对理论的仔细推导远超出了本书的范围，但由于现在许多统计和计量经济软件包都有计算这些统计量的选项，所以对异方差—稳健性方法的应用十分容易。

首先，考虑具有单一自变量的模型，其中我们为强调起见而用了下标i：

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

我们以后统统假定前四个高斯－马尔科夫假定成立。如果误差含有异方差性，那么

$$\mathrm{Var}(u_i \mid x_i) = \sigma_i^2$$

其中我们给 σ^2 加上下标 i，表示误差方差取决于特定 x_i 的值。

将 OLS 估计量写成

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

250 在假定 MLR.1～MLR.4 下（仅没有同方差性假定），并以样本中 x_i 的值为条件，我们可利用与第 2 章同样的论证来证明

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{\mathrm{SST}_x^2} \tag{8.2}$$

其中$\mathrm{SST}_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$ 为 x_i 的总平方和。当对所有的 i 都有 $\sigma_i^2 = \sigma^2$ 时，这个表达式就简化成了通常的形式 σ^2/SST_x。方程（8.2）明确表明，对于简单回归的情形来说，在出现异方差性时，同方差性条件下推导出来的方差公式就不再正确。

由于 $\hat{\beta}_1$ 的标准误直接基于对$\mathrm{Var}(\hat{\beta}_1)$的估计，所以在出现异方差性时，我们就需要一种估计方程（8.2）的方法。怀特（White，1980）说明了这种做法。令 $\hat{u}_i$ 表示原来 y 对 x 做回归所得到的 OLS 残差。那么，对于任何形式的异方差（包括同方差），$\mathrm{Var}(\hat{\beta}_1)$的一个正确估计量都是

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\hat{u}_i^2}{\mathrm{SST}_x^2} \tag{8.3}$$

它很容易从 OLS 回归后的数据中计算出来。

式（8.3）在哪种意义上会是$\mathrm{Var}(\hat{\beta}_1)$的一个正确估计量呢？这相当微妙。简言之，可以证明，将方程（8.3）乘以样本容量 n 后，会按概率收敛于$\mathrm{E}[(x_i - \mu_x)^2u_i^2]/(\sigma_x^2)^2$，即式（8.2）与 n 之积的概率极限。最终，这就是用标准误［指式（8.3）的平方根——译者注］构造置信区间和 t 统计量之所以正确的必要条件。大数定律和中心极限定理在证明这些概率中起到了关键作用。详细内容，可参见怀特的原始论文，但那篇论文相当技术化。也可参见伍德里奇［Wooldridge (1999)，第 4 章］。

在一般多元回归模型

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

中，也有一个类似的公式。可以证明，在假定 MLR.1—MLR.4 下，$\text{Var}(\hat{\beta}_j)$的一个正确估计量是

$$\widehat{\text{Var}}(\hat{\beta}_j)=\frac{\sum_{i=1}^{n}\hat{r}_{ij}^2\hat{u}_i^2}{\text{SSR}_j^2} \tag{8.4}$$

式中，$\hat{r}_{ij}$为将x_j对所有其他自变量做回归所得到的第i个残差；SSR_j则为这个回归的残差平方和（怎样把 OLS 估计值作局部化表示，参见 3.2 节）。式（8.4）的平方根被称为$\hat{\beta}_j$的**异方差—稳健性标准误**（heteroskedasticity-robust standard error）。在计量经济学中，这些稳健性标准误通常都归功于怀特（White，1980）。统计学的更早期著作［比较典型的是 Eicker（1967）和 Huber（1967）］也曾指出了得到这种稳健性标准误的可能性。在应用研究中，有时又把它们称为 White，Huber 或 Eicker 标准误（或用连字符将他们的名字连在一起）。我们只是把它们称为**异方差—稳健性标准误**，甚或在不引起混淆的情况下就称为**稳健性标准误**。

有时，作为对自由度的一种修正，在将式（8.4）开平方之前先乘以$n/(n-k-1)$。进行这种调整的根据是，如果 OLS 残差的平方$\hat{u}_i^2$对所有的观测i都相同——样本中同方差性是最可能的形式，那么我们将得到通常的 OLS 标准误。麦金农和怀特（MacKinnon and White，1985）还研究了对式（8.4）的其他修改。由于所有的形式都只是渐近正确，而且还渐近等价，所以没有哪个形式一定比所有其他的形式都更好。通常，我们总是采用由手边使用的回归软件包所计算出来的随便哪一种形式。

一旦得到了异方差—稳健性标准误，构造一个**异方差—稳健性t统计量**（heteroskedasticity-robust t statistic）就很容易。回想t统计量的一般形式是

$$t=\frac{\text{估计值}-\text{假设值}}{\text{标准误}} \tag{8.5}$$

由于我们仍在使用 OLS 估计值，而且提前选定了假设值，所以通常 OLS 的t统计量和异方差—稳健性t统计量之间惟一的区别是如何计算标准误。

例 8.1 同时使用异方差—稳健性标准误的对数工资方程

虽然我们在例 7.6 中估计了这个模型，但现在我们要在报告通常的 OLS 标准误的同时，也把异方差—稳健性标准误报告出来。有几个估计值在报告时多给出几位数字，以便我们将通常的标准误与异方差—稳健性标准误相比较：

$$\begin{aligned}\widehat{\log(wage)}=\;&\underset{\substack{(0.100)\\ [0.109]}}{0.321}+\underset{\substack{(0.055)\\ [0.057]}}{0.213}\,marrmale-\underset{\substack{(0.058)\\ [0.058]}}{0.198}\,marrfem\\ &-\underset{(0.056)}{0.110}\,singfem+\underset{(0.0067)}{0.0789}\,educ+\underset{(0.0055)}{0.0268}\,exper\end{aligned}$$

$$\begin{array}{llll}
[0.057] & [0.007\,4] & [0.005\,1] & (8.6) \\
-0.000\,54exper^2+0.029\,1tenure & & 0.000\,53tenure^2 & \\
(0.000\,11) & (0.006\,8) & (0.000\,23) & \\
[0.000\,11] & [0.006\,9] & [0.002\,4] & \\
n=526, R^2=0.461 & & &
\end{array}$$

252 在对应的 OLS 估计值下面，通常的 OLS 标准误放在圆括号（ ）中，而异方差—稳健性标准误则放在方括号［ ］中。由于这个方程仍然是由 OLS 估计而来，所以惟一的新内容是方括号中的数字。

从方程（8.6）来看，有几点很明显。首先，在这个特定的应用中，任何一个使用通常的 t 统计量被认为是统计显著的变量，使用异方差—稳健性 t 统计量时仍然是统计显著的。这是因为，这两组标准误相差不大。（相关的 p 值则略有不同，因为稳健性 t 统计量与通常非稳健的 t 统计量不同。）标准误相对变化最大的是 *educ* 的系数：通常的标准误是 0.006 7，而稳健标准误是 0.007 4。不过，这个稳健标准误仍意味着稳健 t 统计量在 10 以上。

方程（8.6）还表明，稳健标准误既可以大于通常的标准误，又可以小于通常的标准误。例如，*exper* 的稳健标准误为 0.005 5，而其通常标准误则为 0.005 1。我们事先并不知道哪个会更大，从经验来看，通常发现稳健标准误比通常的标准误要大。

在结束这个例子之前，我们必须强调，到目前为止，我们还不知道在式（8.6）背后的总体模型中是否出现了异方差性。我们所做的，只是在报告通常的标准误的同时，也报告那些不管是否出现异方差性都正确（渐近地）的标准误。可以看出，在这个例子中，使用稳健标准误没有推翻任何一个重要的结论。虽然这种情况在应用研究中经常发生，但在其他情形下，通常的标准误与稳健标准误之间的差距则大得多。作为这种差别相当明显的例子，参见习题 8.7。

你可能立即就要问如下问题：如果异方差稳健标准误比通常的 OLS 标准误适用的情况更多，那我们为什么还非要使用通常的标准误不可呢？这是一个很恰当的问题。在横截面数据的研究中还用到它们的原因之一是，如果同方差性假定成立，而且误差又服从正态分布，那么，无论样本容量的大小如何（参见第 4 章），通常的 t 统计量都服从精确的 t 分布。而稳健标准误和稳健 t 统计量只有在样本容量越来越大时才能使用。在小样本容量的情况下，稳健 t 统计量的分布可能不是那么接近于 t 分布，从而使我们的推断确实可能犯错误。

在大样本容量的情况下，我们就有理由在横截面数据分析中总是只报告异方差稳健标准误，而且在应用研究中这种做法越来越多。像方程（8.6）那样同时报告两个标准误的做法也很常见，以便读者判断是否有些结论对所使用的标准误有敏感的反应。

还有可能得到任意一个未知形式的对异方差都稳健的 F 和 LM 统计量。

异方差稳健 *F* 统计量（heteroskedasticity-robust *F* statistic）（或其简单变换）又被称为异方差稳健 Wald 统计量。对这个统计量的一般化处理超出了本书的范围。不过，由于许多统计软件包现在都照例计算出这些统计量，所以了解异方差稳健 *F* 和 *LM* 统计量的可用性也有好处。[详细内容可参见 Wooldridge(1999)。] 258

例 8.2　异方差稳健 *F* 统计量

利用 GPA3.RAW 中春季学期的数据，我们估计了如下方程：

$$
\begin{aligned}
\widehat{cumgpa} = {} & 1.47 + 0.00114\,sat - 0.00857\,hsperc + 0.00250\,tothrs \\
& (0.23)(0.00018) \quad (0.00124) \quad (0.00073) \\
& [0.22][0.00019] \quad [0.00140] \quad [0.00073] \\
& + 0.303\,female - 0.128\,black - 0.059\,white \qquad (8.7) \\
& (0.059) \quad (0.147) \quad (0.141) \\
& [0.059] \quad [0.118] \quad [0.110] \\
& n = 366, R^2 = 0.4006, \bar{R}^2 = 0.3905
\end{aligned}
$$

通常的标准误和异方差稳健标准误之间的差别仍不是很大，而且使用稳健 t 统计量不会改变任何一个自变量的统计显著性。联合显著性检验也不会受到太大的影响。假设我们想检验的虚拟假设是在控制了所有其他因素之后，*cumgpa* 在种族之间没有差异。这个虚拟假设可表述为 $H_0: \beta_{black} = 0, \beta_{white} = 0$。一旦我们从受约束模型中得到了 R^2，就很容易得到通常的 *F* 统计量；最后得到这个 R^2 为 0.3983。于是 *F* 统计量为 $[(0.4006 - 0.3983)/(1 - 0.4006)](359/2) \approx 0.69$。如果出现了异方差性，这个检验形式就是不正确的。异方差稳健性的统计量没有简单的形式，但可通过一些特定的统计软件包计算出来。计算出来的异方差稳健性 F 统计量为 0.75，与非稳健形式的数值只是略有不同。这个稳健检验的 p 值为 0.474，与标准的显著性水平并不接近。无论使用哪种检验，我们都不能拒绝虚拟假设。

计算异方差—稳健的 *LM* 检验

并不是所有的回归软件包都计算对异方差稳健的 *F* 统计量。因此，如果能找到一种方法，让我们能得到多重排除性约束的一个异方差稳健性检验，而且又不需要特殊的计量经济软件，那就方便了。那就是任何一个回归软件包都很容易计算的**异方差稳健性 *LM* 统计量**（heteroskedasticity-robust *LM* statistic）。

问题 8.1　评价如下论断：异方差稳健标准误比通常的标准误总是更大。

为了说明稳健 LM 统计量的计算，考虑模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

254 并假设我们想检验$H_0: \beta_4 = 0, \beta_5 = 0$。为了得到通常的 LM 统计量，我们将首先估计受约束模型（即不含 x_4 和 x_5 的模型），以得到残差 $\tilde{u}$。然后再将 $\tilde{u}$ 对所有的自变量进行回归，而且 $LM = n \cdot R_{\tilde{u}}^2$，其中 $R_{\tilde{u}}^2$ 就是从这个回归中得到的 R^2。

要得到异方差稳健形式的统计量还需要更多的工作。有一种只需要 OLS 回归便能计算这种统计量的方法。我们需要将 x_4 对 x_1，x_2，x_3 做回归所得到的残差 $\tilde{r}_1$ 和将 x_5 对 x_1，x_2，x_3 做回归所得到的残差 $\tilde{r}_2$。于是，我们将虚拟假设中所排除的自变量对虚拟假设中所包括的所有自变量做回归，并且每次都得到一个残差。最后一步看上去有些古怪（但它毕竟只是一个计算工具），即做如下不包括截距项的回归

$$1 \text{ 对 } \tilde{r}_1 \tilde{u}, \tilde{r}_2 \tilde{u} \text{ 回归} \tag{8.8}$$

是的，我们实际上定义了一个对所有观测都等于 1 的因变量。我们将这个因变量对乘积 $\tilde{r}_1 \tilde{u}$ 和 $\tilde{r}_2 \tilde{u}$ 做回归。稳健性 LM 统计量其实就是$n - \mathrm{SSR}_1$，其中 SSR_1 刚好是回归（8.8）中通常的残差平方和。

究其原因，多少有些技术化。基本上看来，这是在构造适合于 LM 检验的稳健标准误，如同构造适合于 t 检验的稳健标准误那样。［更详尽的讨论，可参见 Wooldridge（1991b）或 Davidson and MacKinnon（1993）。］

现在，我们在一般情形下总结对异方差稳健性 LM 统计量的计算。

异方差稳健性 *LM* 统计量

1. 从受约束模型中得到。

2. 将虚拟假设中排除的每个自变量对虚拟假设所包括的所有自变量做回归；如果有 q 个被排除变量，那就得到 q 个残差$(\tilde{r}_1, \tilde{r}_2, \cdots, \tilde{r}_q)$构成的集合。

3. （对所有的观测都）求出每个 $\tilde{r}_j$ 和 $\tilde{u}$ 的积。

4. 在不包括截距的情况下将 1 对 $\tilde{r}_1 \tilde{u}$，$\tilde{r}_2 \tilde{u}$，…，$\tilde{r}_q \tilde{u}$ 做回归。异方差稳健性的 LM 统计量就是$n - \mathrm{SSR}_1$，其中 SSR_1 就是最后这个回归的通常残差平方和。在 H_0 下，LM 渐近服从分布 χ_q^2。一旦得到了稳健性 LM 统计量，假设的拒绝规则和 p 值的计算都与 5.2 节中通常的 LM 统计量一样。

例 8.3　异方差稳健性 *LM* 统计量

255 我们利用 CRIME1.RAW 中的数据，来检验过去在定罪后的平均判刑期对当年（1986 年）被拘捕的次数有没有影响。所估计的模型是

$$\underset{}{narr86} = \underset{(0.036)}{0.567} - \underset{(0.040)}{0.136}\,pcnv + \underset{(0.009\,7)}{0.017\,8}\,avgsen - \underset{(0.000\,30)}{0.000\,52}\,avgsen^2$$

$$
\begin{array}{llll}
[0.040][0.034] & [0.0101] & [0.000\,21] & \\
-0.394ptime86 & -0.050\,5qemp86 & -0.001\,48inc86 & \\
(0.008\,7) & (0.014\,4) & (0.000\,34) & \\
[0.006\,2] & [0.014\,2] & [0.000\,23] & \\
+0.325black & +0.193hispan & & (8.9) \\
(0.045) & (0.040) & & \\
[0.058] & [0.040] & & \\
n=2\,725, R^2=0.072\,8 & & &
\end{array}
$$

在这个例子中，某些通常的标准误和稳健标准误之间存在着更大的差别。比如对变量 $avgsen^2$ 来说，通常的 t 统计量约为 1.73，而稳健的 t 统计量约为 -2.48。因此，$avgsen^2$ 在使用稳健标准误时更显著。

avgsen 对 *narr*86 的影响多少有些难以解释。由于二者之间存在着二次关系，所以我们能计算出什么时候对 *narr*86 有正影响，什么时候对 *narr*86 有负影响。转折点是 $0.017\,8/[2(0.000\,52)]\approx 17.12$；记住它是按月度量的。这确实意味着，当 *avgsen* 低于 17 个月时，*narr*86 与 *avgsen* 就正相关；而在 17 个月以后，预期 *avgsen* 会具有威慑作用。

为了看出平均判刑期对 *narr*86 是否具有统计显著的影响，我们必须检验联合假设 $\mathrm{H}_0: \beta_{avgsen}=0, \beta_{avgsen^2}=0$。利用通常的 LM 统计量（参见 5.2 节），我们得到 $LM=3.54$；在一个自由度为 2 的 χ 平方分布中，可得到一个 p 值 $=0.170$。于是，我们即使在 15% 的显著性水平上都不能拒绝 H_0。异方差稳健的 LM 统计量为 $LM=4.00$（保留两位小数），p 值 -0.135。这仍不是拒绝 H_0 的很强的证据；看来 *avgsen* 对 *narr*86 没有一种很强的影响。[随便指出，当 *avgsen* 单独出现在式(8.9)时，也就是说没有二次项，通常的 t 统计量为 0.658，而其稳健 t 统计量为 0.592。]

8.3 对异方差性的检验

无论是否存在异方差性，异方差—稳健标准误都为计算渐近于 t 分布的 t 统计量提供了一种简单方法。我们还看到，异方差稳健的 F 和 LM 统计量都是现成的。进行这些检验时并不需要知道是否存在异方差性。不过，仍有一些很好的理由，要求找到一些能侦查其是否存在的简单检验。首先，正如我们在上一节中所提到的那样，通常的 t 统计量在经典线性模型假定之下具有精确的 t 分布。为此，许多经济学家仍然更希望看到、报告的是通常的 OLS 标准误和检验统计量，除非存在有异方差性的证据。其次，如果存在异方差性，那么 OLS 估计量就不再是最优线性无偏估计量。如我们在 8.4 节中将看到的那样，当异方差的形式已知时，有可能会得到一个比 OLS 更好的估计量。

多年来，人们已经提出过许多种检验异方差性的方法。其中有些方法尽

管有能力侦查异方差性，但并不直接检验误差方差与自变量无关的假定。我们将仅考虑较现代的检验，它们能侦查出使通常的OLS统计量无效的异方差类型，同时还有将所有检验都放在同一框架之中的好处。

像平常一样，我们从如下线性模型开始：

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k+u \tag{8.10}$$

在本节仍然维持假定MLR.1～MLR.4，特别是我们假定$\mathrm{E}(u|x_1,x_2,\cdots,x_k)=0$，从而OLS是无偏而又一致的。

我们取虚拟假设为：假定MLR.5是正确的，即

$$\mathrm{H}_0:\mathrm{Var}(u|x_1,x_2,\cdots,x_k)=\sigma^2 \tag{8.11}$$

即同方差性这个理想的假定成立，并要求在不成立时，数据能够告诉我们。如果我们在一个充分小的显著性水平上不能拒绝式（8.11），那么我们通常就会断定没有异方差性的问题。但请记住，我们绝不会接受H_0；我们只不过是不能拒绝而已。

由于我们假定u的条件期望值为零，所以$\mathrm{Var}(u|\boldsymbol{x})=\mathrm{E}(u^2|\boldsymbol{x})$，因而同方差性的虚拟假设就等价于

$$\mathrm{H}_0:E(u^2|x_1,x_2,\cdots,x_k)=E(u^2)=\sigma^2$$

这说明，为了检验真实情况是否违背了同方差性假定，我们想检验u^2是否与一个或多个解释变量相关（在期望值的意义上）。如果H_0是错误的，那么给定自变量，u^2的期望值就可能是x_j的某个函数。一个简单的方法就是假定一个线性函数：

$$u^2=\delta_0+\delta_1x_1+\delta_2x_2+\cdots+\delta_kx_k+v \tag{8.12}$$

式中，v为给定x_j下均值为零的一个误差项。仔细看看这个方程中的因变量：它是原回归方程（8.10）中误差项的平方。同方差性的虚拟假设是

$$\mathrm{H}_0:\delta_1=\delta_2=\cdots=\delta_k=0 \tag{8.13}$$

在这个虚拟假设下，常常有理由假定式（8.12）中的误差v与x_1，x_2，…，x_k无关。然后我们从5.2节知道，为检验解释u^2的自变量的整体显著性，可用F或LM统计量来检验式（8.13）。尽管u^2不是正态分布的（比如，如果u是正态分布的，那么u^2/σ^2就服从χ_1^2分布。），但这两个统计量都是渐近正确的。如果我们能观测到样本中的u^2，那么利用所有n个观测，通过u^2对x_1，x_2，…，x_k的OLS回归，就能轻而易举地计算出这个统计量。

257 如前面曾强调过的那样，我们虽然永远不知道总体模型中的实际误差，但确实能得到它们的估计值：OLS残差$\hat{u}_i$是第i个观测的误差u_i的一个估计值。因此，我们可以估计方程

$$\hat{u}^2=\delta_0+\delta_1x_1+\delta_2x_2+\cdots+\delta_kx_k+误差 \tag{8.14}$$

并对x_1，x_2，…，x_k的联合假设计算F或LM统计量。结果表明，用OLS

残差取代误差并不影响 F 或 LM 统计量的大样本分布性质，尽管相当复杂。

F 和 LM 统计量都取决于回归式（8.14）的 R-平方；为与估计方程（8.10）所得到的 R-平方相区别，我们把它称为 $R^2_{\hat{u}^2}$。于是，F 统计量就是

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n - k - 1)} \tag{8.15}$$

式中，k 为式（8.14）中回归元的个数，与式（8.10）中自变量的个数相同。由于大多数回归软件包都计算检验回归整体显著性的 F 统计量，所以几乎没有必要手算式（8.15）。在同方差性的虚拟假设之下，这个 F 统计量（渐近地）服从一个 $F_{k,n-k-1}$ 分布。

同方差性的 LM 统计量恰好是样本容量乘以式（8.14）的 R-平方：

$$LM = n \cdot R^2_{\hat{u}^2} \tag{8.16}$$

在虚拟假设之下，LM 渐近服从 χ^2_k 分布。在做了回归式（8.14）后，这个统计量也很容易得到。

这个 LM 形式的检验被典型地称为 **Breusch-Pagan 异方差检验**（Breusch Pagan test for heteroskedasticity, BP test），布罗施和培甘（Breusch and Pagan, 1980）还提出了一个假定误差正态分布的检验形式。考恩克（Koenker, 1983）也提出了式（8.16）中 LM 统计量的检验形式，由于其更大的适用性，所以一般更受欢迎。

我们把用 BP 检验来检验异方差性的步骤总结如下。

异方差性的 Breusch-Pagan 检验

1. 像平常一样用 OLS 估计模型（8.10）。得到 OLS 残差平方 $\hat{u}^2$（每次观测得到一个）。

2. 做式（8.14）中的回归。记下这个回归的 R-平方 $R^2_{\hat{u}^2}$。

3. 计算 F 统计量或 LM 统计量并计算 p 值（前者用 $F_{k,n-k-1}$ 分布，后者用 χ^2_k 分布）。如果这个 p 值相当小，即低于选定的显著性水平，那么我们就拒绝同方差性的虚拟假设。

258 如果 BP 检验得到一个足够小的 p 值，那就应该采取某种校正措施。一种可能措施就是使用异方差稳健的标准误，并检验上一节中讨论的统计量。另一种可能措施在 8.4 节讨论。

例 8.4　住房价格方程中的异方差性

我们使用 HPRICE1.RAW 中的数据来检验一个简单的住房价格方程中的异方差性。利用所有变量的水平值所估计的方程是

$$\begin{aligned}\widehat{price} &= -21.77 + 0.002\,07 lotsize + 0.123 sqrft + 13.85 bdrms \\ &\quad\ (29.48)\quad (0.000\,64) \qquad\quad (0.013) \qquad\ (9.01) \\ &\quad\ n = 88, R^2 = 0.672\end{aligned} \tag{8.17}$$

这个方程丝毫也没有告诉我们此总体模型中的误差是否存在异方差性。我们需要将 OLS 残差的平方对自变量做回归。$\hat{u}^2$ 对 *lotsize*，*sqrft* 和 *bdrms* 回归所得到的 R-平方为 $R_{\hat{u}^2}^2 = 0.1601$。在 $n = 88$ 和 $k = 3$ 的情况下，得到检验自变量显著性的 F 统计量 $F = [0.1601/(1-0.1601)](84/3) \approx 5.34$。相应的 p 值为 0.002，它是拒绝虚拟假设的有力证据。LM 统计量为 $88(0.1601) \approx 14.09$；它给出的 p 值 ≈ 0.0028（利用 χ_3^2 分布），从而得到与 F 统计量本质上同样的结论。这就意味着式(8.17)中报告的通常的标准误是不可靠的。

我们在第 6 章中提到，使用因变量的对数函数形式有一个好处，就是通常能够消除异方差性。在本例中，我们取 *price*，*lotsize* 和 *sqrft* 的对数形式，使得 *price* 对 *lotsize* 和 *sqrft* 的弹性为常数。所估计的方程是

$$
\begin{aligned}
\log(price) = {} & 5.61 + 0.168\log(lotsize) + 0.700\log(sqrft) \\
& (0.65)\;(0.038) \qquad\qquad\quad (0.093) \\
& + 0.037\,bdrms \\
& (0.028) \qquad\qquad\qquad\qquad\qquad\qquad (8.18) \\
& n = 88,\ R^2 = 0.643
\end{aligned}
$$

将这个回归的 OLS 残差平方对 $\log(lotsize)$，$\log(sqrft)$ 和 *bdrms* 做回归，给出 $R_{\hat{u}^2}^2 = 0.0480$。因此，$F = 1.41$（p 值 $= 0.245$），而 $LM = 4.22$（p 值 $= 0.239$）。所以，我们不能拒绝对数函数形式模型中同方差性的虚拟假设。我们注意到，在许多经验研究中都出现过对数形式的因变量的情况，其异方差性较弱。

如果我们猜测异方差性只取决于某些自变量，那我们很容易就能改造 Breusch-Pagan 检验：只要将 $\hat{u}^2$ 对我们所选择的不管哪些自变量做回归，并进行适当的 F 或 LM 检验。记住，恰当的自由度取决于以 $\hat{u}^2$ 为因变量的回归中自变量的个数；方程（8.10）中出现的自变量的个数并不重要。

如果残差平方只对单一个自变量做回归，那么异方差检验恰好就是该变量通常的 t 统计量。显著的 t 统计量表示异方差性是个问题。

问题 8.2 ☞

考虑工资方程（7.11），并且你认为 $\log(wage)$ 与其中的 *educ*，*exper* 或 *tenre* 都无关，不过，你担心 $\log(wage)$ 的方差在已婚男性、已婚女性、单身男性和单身女性这四个人口组之间不同。为了检验异方差性，你应该做什么样的回归？F 检验的自由度是多少？

异方差性的White检验

250　我们在第 5 章证明了，如果所有的高斯-马尔科夫假定都成立的话，通

常的 OLS 标准误和检验统计量都是渐近生效的。此时，同方差性假定 $\mathrm{Var}(u_1|x_1,\cdots,x_k)=\sigma^2$ 可由如下较弱的假定所取代，即误差平方 u^2 与所有自变量（x_j）、所有自变量的平方（x_j^2）和所有自变量的交叉乘积（$x_jx_h, j\neq h$）都不相关。这一观察，促使怀特（White，1980）提出对异方差性的一种检验方法，即在方程（8.14）中增加所有自变量的平方和交叉乘积项。这个检验明显有意用于检验那些使通常的 OLS 标准误和检验统计量无效的异方差形式。

当模型包含 $k=3$ 个自变量时，White 检验则基于如下估计：

$$\begin{aligned}\hat{u}^2 &= \delta_0+\delta_1x_1+\delta_2x_2+\delta_3x_3+\delta_4x_1^2+\delta_5x_2^2\\&\quad+\delta_6x_3^2+\delta_7x_1x_2+\delta_7x_1x_3+\delta_9x_2x_3+\text{误差项}\end{aligned}\tag{8.19}$$

与 Breusch-Pagan 检验相比，这个方程多了 6 个回归元。**White 异方差检验**（White test for heteroskedasticity），就是检验方程（8.19）中除截距外所有的 δ_j 都为 0 的 LM 统计量。因而在这个情形下，要检验 9 个约束。对于这个假设，我们也可以使用 F 检验；这两个检验都具有渐近合理性。

在原方程只有 3 个自变量的情况下，方程（8.19）就有 9 个自变量。原方程若有 6 个自变量，White 回归一般会涉及 27 个回归元（除非某些是多余的）。回归元过多是纯粹形式 White 检验的一个缺陷：对于那些自变量个数适中的模型，它要使用掉很多自由度。

有可能得到一个比 White 检验更容易实施而且自由度更节省的检验。为了得到这个检验，回忆 White 检验与 Breusch-Pagan 检验之间的差别，前者包括了自变量的平方项和交叉乘积项。我们不需用那么多自变量的函数，同样可以做到这一点。一种建议是，在异方差检验中使用 OLS 拟合值。记住，对于每次观测 i，拟合值都被定义为

$$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_{i1}+\hat{\beta}_2x_{i2}+\cdots+\hat{\beta}_kx_{ik}$$

它们只不过是自变量的线性函数。如果我们将拟合值平方，那么就得到自变量所有平方项和所有交叉乘积项的一个特殊函数。这就表明，通过估计方程

260

$$\hat{u}^2=\delta_0+\delta_1\hat{y}+\delta_2\hat{y}^2+\text{误差项}\tag{8.20}$$

式中，$\hat{y}$ 为拟合值，就能检验异方差性。重要的是，在这个方程中不要将 $\hat{y}$ 和 y 相混淆。我们使用拟合值是因为它们是自变量（及所估计参数）的函数；在式（8.20）中使用 y，不会得到异方差性的有效检验。

对于方程（8.20）中的虚拟假设 $\mathrm{H}_0:\delta_1=0$ 和 $\delta_2=0$，我们可以使用 F 或 LM 统计量。这样就导致了无论原模型中有多少个自变量，检验同方差性的虚拟假设时都只有两个约束。如此减少自由度常常是个好办法，而且还使检验易于实施。

由于在给定 x_j 时，$\hat{y}$ 是对 y 的期望值的一个估计，所以在认为方差随着期望值 $\mathrm{E}(y|\boldsymbol{x})$ 而变化时，利用式（8.20）来检验异方差性就很有用。因为可以把方程（8.20）看成对方程（8.19）中的参数施加了限制，所以，式（8.20）中的检验就可以看成是 White 检验的一种特殊情形。

White 异方差检验的特例

1. 像平常一样用 OLS 估计模型（8.10），得到 OLS 残差 $\hat{u}$ 和拟合值 $\hat{y}$。计算 OLS 残差的平方 $\hat{u}^2$ 和拟合值的平方 $\hat{y}^2$。

2. 做方程（8.20）中的回归。记下这个回归的 R-平方 $R^2_{\hat{u}^2}$。

3. 构造 F 或 LM 统计量并计算 p 值（前者用 $F_{2,n-3}$ 分布，后者用 χ^2_2 分布）。

例 8.5　对数住房价格方程中 White 检验的特殊形式

我们将 White 检验的特殊情形应用到方程（8.18）中，其中我们使用 LM 形式的统计量。要记住的重要一点是，χ^2 分布的自由度总是 2。$\hat{u}^2$ 对 $lprice$，$(lprice)^2$ 的回归［其中 $lprice$ 表示从式（8.18）得到的拟合值］给出 $R^2_{\hat{u}^2}=0.0392$；于是，$LM=88(0.0392)\approx3.45$，而 p 值 $=0.178$。与 Breusch-Pagan 检验所提供的结论相比，这是存在异方差性的更强的证据，但我们即使在 15% 的显著性水平上仍不能拒绝同方差性的虚拟假设。

在结束本节之前，我们应该讨论一个重要的解释。我们把利用一个异方差检验而拒绝虚拟假设解释为存在异方差性的证据。如果我们维持假定 MLR.1～MLR.4，这种解释就是适当的。但如果违背了 MLR.3［具体而言，就是 E(y|x)的函数形式被错误地设定］，那么，即使 Var(y|x)是常数，异方差检验也可能会拒绝 H_0。比如，如果我们在一个回归模型中漏掉了一个或多个二次项，或者在应该使用对数模型时却使用了水平模型，对异方差性的检验都可能是显著的。这一点已经使某些经济学家将异方差检验看成一般的模型误设检验。261 不过，对于函数形式误设的问题，有更好、更直接的检验，我们将在 9.1 节讨论一些这样的检验。既然函数形式误设的问题比异方差性更重要，所以最好是先对函数形式进行明确的检验。然后，如果我们对函数形式感到满意了，就可以检验异方差性。

8.4　加权最小二乘估计

如果利用 8.3 节中的检验方法之一发现存在异方差性，那么，我们在 8.2 节了解到，一种可能的回应是，在用 OLS 估计之后使用异方差稳健的统计量。在提出异方差稳健统计量之前，对发现存在异方差性的回应是，建立模型并估计其具体形式。如我们将看到的那样，这样会得到一个比 OLS 更有效的估计量，而且还由此得到具有 t 和 F 分布的 t 和 F 统计量。尽管这看上去颇具吸引力，但因为我们必须对任何一种异方差的性质十分明确，所以，我们实际上还有很多工作要做。

除一个常数倍数外异方差是已知的

令 $\boldsymbol{x}$ 表示方程（8.10）中所有的解释变量，并假定

$$\mathrm{Var}(u|\boldsymbol{x})=\sigma^2h(\boldsymbol{x}) \tag{8.21}$$

式中，$h(\boldsymbol{x})$为解释变量的某种函数，并决定着异方差性。由于方差必须为正，所以对所有可能的自变量值都有$h(\boldsymbol{x})>0$。我们在本小节假定函数$h(\boldsymbol{x})$为已知。虽然总体参数 σ^2 未知，但我们能把它从一个数据样本中估计出来。

对于从总体中的一个随机抽取，我们可以写出$\sigma_i^2=\mathrm{Var}(u_i|\boldsymbol{x}_i)=\sigma^2h(\boldsymbol{x}_i)=\sigma^2h_i$，其中我们再次用 $\boldsymbol{x}_i$ 表示第 i 次观测所有的自变量，而由于自变量随着观测而变化，所以 h_i 随着每次观测而变化。比如，考虑简单的储蓄函数（8.22）：

$$sav_i=\beta_0+\beta_1 inc_i+u_i \tag{8.22}$$

$$\mathrm{Var}(u_i|inc_i)=\sigma^2 inc_i \tag{8.23}$$

这里，$h(inc)=inc$：误差方差与收入水平成正比。这意味着，随着收入的提高，储蓄的可变性也在提高。（如果 $\beta_1>0$，储蓄的期望值也随着收入的提高而增加。）由于 inc 总是正的，所以方程（8.23）中的方差总能保证为正。u_i 的标准差（以 inc_i 为条件）就是 $\sigma\sqrt{inc_i}$。

我们如何用方程（8.21）中的信息去估计 β_j 呢？本质上，我们先取包含了异方差误差的原方程

262

$$y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_kx_{ik}+u_i \tag{8.24}$$

然后把它转换成一个具有同方差误差的方程（并满足其他的高斯-马尔科夫假定）。由于 h_i 仅是 $\boldsymbol{x}_i$ 的函数，所以$u_i/\sqrt{h_i}$以 $\boldsymbol{x}_i$ 为条件的期望值为零。而且，由于$\mathrm{Var}(u_i|\boldsymbol{x}_i)=\mathrm{E}(u_i^2|\boldsymbol{x}_i)=\sigma^2h_i$，所以$u_i/\sqrt{h_i}$的方差（以 $\boldsymbol{x}_i$ 为条件）为 σ^2：

$$\mathrm{E}((u_i/\sqrt{h_i})^2)=\mathrm{E}(u_i^2)/h_i=(\sigma^2h_i)/h_i=\sigma^2$$

其中为简单起见我们已经省略了以 $\boldsymbol{x}_i$ 为条件。我们可以将方程（8.24）两边同时除以$\sqrt{h_i}$而得到

$$\begin{aligned}y_i/\sqrt{h_i}={}&\beta_0/\sqrt{h_i}+\beta_1(x_{i1}/\sqrt{h_i})+\beta_2(x_{i2}/\sqrt{h_i})+\cdots+\beta_k(x_{ik}/\sqrt{h_i})\\&+(u_i/\sqrt{h_i})\end{aligned} \tag{8.25}$$

或

$$y_i^*=\beta_0x_{i0}^*+\beta_1x_{i1}^*+\cdots+\beta_kx_{ik}^*+u_i^* \tag{8.26}$$

式中，$x_{i0}^*=1/\sqrt{h_i}$，其他标星号的变量都表示将对应原变量除以$\sqrt{h_i}$。

方程（8.26）看起来有些奇怪，但重要的是要记住，我们之所以这样推

导，是为了得到比 OLS 的效率性质更好的 β_j 的估计量。原方程（8.24）中的截距项 β_0 现在被乘上了变量 $x_{i0}^* = 1/\sqrt{h_i}$。β_j 中的每一个斜率参数都乘上了一个新的变量，使得对这些新变量很难作出什么有用的解释。但如果我们联想到，为了解释模型及其参数，总想回到原方程（8.24），这样做就不应该引起什么问题。

在前面储蓄方程的例子中，变形后的方程为

$$sav_i/\sqrt{inc_i} = \beta_0(1/\sqrt{inc_i}) + \beta_1\sqrt{inc_i} + u_i^*$$

式中，我们利用了等式 $inc_i/\sqrt{inc_i} = \sqrt{inc_i}$。不过，$\beta_1$ 仍是收入的边际储蓄倾向，这是我们从方程（8.22）得到的解释。

方程（8.26）对其参数而言是线性的（所以 MLR.1），而且随机抽样的假定也没有改变。此外，以 x_i^* 为条件，u_i^* 具有零均值和常方差（σ^2）。这就意味着，如果原方程满足了前四个高斯-马尔科夫假定，那么，变换后的方程（8.26）就满足所有的 5 个高斯-马尔科夫假定。同时，如果 u_i 具有正态分布，那么 u_i^* 也具有方差为 σ^2 的正态分布。因此，如果原模型满足除同方差性假定外所有的经典线性模型假定，那么变换后的方程就会满足全部的经典线性模型假定（MLR.1～MLR.6）。

因为我们知道 OLS 在高斯-马尔科夫假定下具有诱人的性质（比如 BLUE），所以上一段的讨论建议我们用普通最小二乘法来估计方程（8.26）中的参数。这些估计量 β_0^*，β_1^*，…，β_k^* 将与原方程中的 OLS 估计量有所

263 区别。这些 β_j^* 正是**广义最小二乘（GLS）估计量**（generalized least squares estimators）的例子。这里，GLS 估计量被用来对付误差中的异方差性。我们在第 12 章还将碰到其他的 GLS 估计量。

由于方程（8.26）满足所有的理想假定，所以从利用变换后的变量所做的回归中，能够得到标准误、t 统计量和 F 统计量。式（8.26）的残差平方和除以自由度就是 σ^2 的一个无偏估计量。而且，因为这些 GLS 估计量都是 β_j 的最优线性无偏估计量，所以必然比从原方程得到的 OLS 估计量 $\hat{\beta}_j$ 更有效。实质上，在将变量变换之后，我们就只需要进行标准的 OLS 分析。但我们必须记住，要将估计值放到原方程中去解释。

估计式（8.26）所得到的 R-平方，尽管对计算 F 统计量很有用，但作为拟合优度的指标没有什么特别的含义：它告诉我们 y^* 的变异中有多少能由 x_j^* 来解释，但这几乎没有什么意义。

这种纠正异方差性的 GLS 估计量又被称为**加权最小二乘（WLS）估计量**（weighted least squares estimators）。这个名称来自如下事实：β_j^* 最小化了残差平方的加权和，其中每个残差平方的权数都为 $1/h_i$。其思想是，对误差方差越大的观测赋予越小的权数；OLS 则对每个观测都赋予相同的权数，因为在总体的每部分的误差方差都相同时，这样做是最好的。从数学上讲，WLS 估计量就是使下式尽可能小的 b_j 值：

$$\sum_{i=1}^{n}(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - \cdots - b_kx_{ik})^2/h_i \qquad (8.27)$$

将 $1/h_i$ 的平方根放进残差平方的表达式里，表明加权后的残差平方和等于变换后变量的残差平方和：

$$\sum_{i=1}^{n}(y_i^* - b_0x_{i0}^* - b_1x_{i1}^* - b_2x_{i2}^* - \cdots - b_kx_{ik}^*)^2$$

于是，最小化式（8.27）的 WLS 估计量无非就是从式（8.26）得到的 OLS 估计量。

加权最小二乘估计量可以对任何一组正的权数加以定义，而 OLS 正是对所有观测都赋予相等权数的特殊情形。有效的 GLS 程序赋予每个残差平方的权数都是 u_i 在给定 $\mathbf{x}_i$ 时的条件方差的倒数。

为进行加权最小二乘而对变量进行变形可能很繁冗，而且难免会出差错。幸运的是，大多数现代回归软件包都能做加权最小二乘估计。通常，在给出原模型中自变量和因变量的同时，只须确定加权函数即可。这样做不仅不太容易出错，而且好让我们能在原模型中来解释加权最小二乘估计值。实际上，我们可以用通常的方式写出所估计的方程。虽然估计值和标准误与 OLS 所得到的不同，但我们解释那些估计值、标准误和检验统计量的方式都是一样的。

例 8.6　家庭储蓄方程

264 表 8.1 包含了得自数据集 SAVING.RAW（1970 年以来的 100 个家庭）的储蓄方程的估计值。我们用 OLS 和加权最小二乘法分别估计了简单回归模型式（8.22），并假定后者的方差由式（8.23）给出。然后我们添加变量如家庭规模、户主年龄、户主受教育年数和一个表明户主是否黑人的虚拟变量。

在简单回归模型中，边际储蓄倾向（MPS）的 OLS 估计值为 0.147，t 统计量为 2.53。（表 8.1 中 OLS 的标准误是非稳健标准误。如果确实认为异方差性成问题，那我们可能还需要计算异方差—稳健标准误；我们在这里不予计算。）MPS 的 WLS 估计值多少高一些：0.172，$t=3.02$。这一系数的 OLS 和 WLS 估计值的标准误十分相似。虽然 OLS 和 GLS 的截距估计值相去甚远，但由于其 t 统计量都很小，所以不必太关注。在将 OLS 和 WLS 估计值进行比较时，发现系数有相当大的差距而又不显著的并非少见。第（1）列和第（2）列中的 R-平方是不可比的。

无论使用 OLS 还是 WLS，增加人口统计数据方面的变量都会降低 MPS；标准误也有相当大的提高（因为增加这些附加变量所导致的多重共线性）。很容易看出，无论使用 OLS 还是 WLS 估计值，这些附加变量中没有一个是个别显著的。那它们是联合显著的吗？基于 OLS 估计值的 F 检验使用第(1)列和第(3)列中的 R-平方。在不受约束模型的自由度为 94 和约束个数为 4 的情况下，F 统计量为 $F=[(0.0828-0.0621)/(1-0.0828)](94/4)\approx0.53$，而 p 值 $=0.715$。利用 WLS 估计值的 F 检验使用第(2)列和第(4)列中的 R-平方：$F\approx0.5$，而 p 值 $=0.739$。因此，无论是用 OLS 估计值

还是 WLS 估计值，人口统计变量都不是联合显著的。这表明，将储蓄与收入相联系的简单回归模型就足够了。

表 8.1 **因变量：*sav***

自变量	(1) OLS	(2) WLS	(3) OLS	(4) WLS
inc	0.147 (0.058)	0.172 (0.057)	0.109 (0.071)	0.101 (0.077)
size	—	—	67.66 (222.96)	6.87 (168.43)
educ	—	—	151.82 (117.25)	139.48 (100.54)
age	—	—	0.286 (50.031)	21.75 (41.31)
black	—	—	518.39 (1 308.06)	137.28 (844.59)
截距	124.84 (655.39)	−124.95 (480.86)	−1 605.42 (2 830.71)	−1 854.81 (2 351.80)
观测次数	100	100	100	100
R-平方	0.062 1	0.085 3	0.082 8	0.104 2

那我们究竟应该选择哪个作为边际储蓄倾向的最佳估计值呢？在这里的情况下，无论我们选择 OLS 估计值 0.147 还是 WLS 估计值 0.172，都没有 265 什么问题。记住，二者都只是从一个相对小的样本中得到的估计值，而且 OLS 估计值 95% 的置信区间也包含了 WLS 估计值，WLS 估计值 95% 的置信区间也包含了 OLS 估计值。

实践中，我们很少知道方差是如何以一个简单的形式取决于某个特定自变量的。比如，在包括了所有人口统计变量的储蓄方程中，我们怎么会知道 *sav* 的方差不会随着年龄或受教育水平而变化呢？在多数应用研究中，我们都不能确定 $\mathrm{Var}(y|x_1, x_2, \cdots, x_k)$。

问题 8.3 ☞

利用表 8.1 中第(1)列所报告的 OLS 回归中得到的残差，$\hat{u}^2$ 对 *inc* 回归得到 *inc* 的 t 统计量为 0.96。有使用例 8.6 中加权最小二乘法的必要吗？

有一种情况，WLS 所需要的权数会自然来自其潜在的计量模型。如果我们拥有的不是个人水平上的数据（指每个人的个别数据——译者注），而是某个组或某个地理区域中的数据的平均值，就会发生这种情况。比如，假使我们对决定一个工人对其 401k 养老金计划参与数额与该计划受到的慷慨

捐献之间的函数关系感兴趣，令 i 表示一个特定的企业，而 e 表示该企业内的某一个雇员。一个简单的模型是

$$contrib_{i,e}=\beta_0+\beta_1 earns_{i,e}+\beta_2 age_{i,e}+\beta_3 mrate_i+u_{i,e} \qquad (8.28)$$

式中，$contrib_{i,e}$为第 i 个企业工作的雇员 e 每年参与的数额；$earns_{i,e}$为此人每年的收入；$age_{i,e}$为此人的年龄；$mrate_i$ 为该企业将雇员贡献的每一美元中放进其账户的数量。

如果式（8.28）满足高斯-马尔科夫假定，那么，给定不同雇主所雇用的雇员样本，我们就能估计它。不过，假如我们只有雇主提供的贡献值、收入和雇员年龄的平均值。换句话说，没有个人水平的数据可供使用，令 $\overline{contrib}_i$ 表示第 i 个企业中雇员的平均贡献，$\overline{earns}_i$ 和$\overline{age}_i$ 也都做类似定义。令 m_i 表示企业 i 的雇员人数；我们假定这是一个已知的数量。于是，如果我们将方程（8.28）对第 i 个企业中所有的雇员进行平均，就能得到企业水平上的方程

$$\overline{contrib}_i=\beta_0+\beta_1\overline{earns}_i+\beta_2\overline{age}_i+\beta_3 mrate_i+\bar{u}_i \qquad (8.29)$$

266 式中，$\bar{u}_i=m_i^{-1}\sum_{e=1}^{m_i}u_{i,e}$，表示第 i 个企业中所有雇员的平均误差。如果我们的样本中有 n 个企业，那么式（8.29）刚好就是一个可由 OLS 来估计的标准多元线性回归模型。如果原模型（8.28）满足高斯-马尔科夫假定，这个估计量就是无偏的，而个人的误差都与企业规模无关［因为这样一来给定式（8.29）中的解释变量，$\bar{u}_i$ 的期望值为零］。

如果个人水平上的方程满足同方差性假定，那么在企业水平上的方程(8.29)就一定是异方差的。事实上，如果对所有的 i 和 e 都有$\mathrm{Var}(u_{i,e})=\sigma^2$，那么就有 $\mathrm{Var}(\bar{u})_i=\sigma^2/m_i$。换句话说，企业越大，其误差项 $\bar{u}_i$ 的方差就会随着企业规模的扩大而减小。在这种情况下，$h_i=1/m_i$，因而最有效的估计程序就是以企业雇员人数为权数($1/h_i=m_i$)的加权最小二乘估计。这就保证了较大的企业得到较大的权数。从而为我们在只有企业水平的平均值的情况下，给出估计个人水平模型中的参数的一种有效方法。

在使用城市、县、州或国家水平的人均数据时，也会出现类似的加权。如果个人水平的方程满足高斯-马尔科夫假定，那么，人均方程中的误差方差就与人口大小的倒数成比例。因此，以人口大小为权数的加权最小二乘法就比较合适。比如，假设有城市水平上的人均啤酒消费量（以盎司为单位）、总人口中年龄在 21 岁以上的人口百分比、成年人的平均受教育水平、平均收入水平和城市啤酒价格水平方面的数据，那么，以城市人口为权数的加权最小二乘法就能估计如下城市水平模型：

$$beerpc=\beta_0+\beta_1 perc21+\beta_2 avgeduc+\beta_3 incpc+\beta_4 price+u^*$$

以企业规模、城市人口为权数的优势，必须以其背后个人方程的同方差性为

---

* 原文的斜率参数出现排版错误。——译者注

条件。如果个人水平的方程中存在着异方差性，那么恰当的权数就取决于异方差的形式。这正是为什么越来越多的研究者在估计人均数据模型时也简单地计算稳健的标准误和检验统计量的原因之一。另一种办法是，以人口进行加权，但在 WLS 估计中报告异方差稳健的统计量。这就确保了在个人水平的模型满足高斯-马尔科夫假定时，估计是有效的；又通过稳健的推断对个人水平模型中任何形式的异方差做了考虑。

必须估计异方差函数：可行GLS

在上一小节中，我们看到了几个异方差已知为乘积形式的例子。在大多数情况下，异方差的确切形式并不明显。换句话说，很难找到上一节中的函数 $h(\mathbf{x}_i)$。不过，在多数情况下，我们可以建造函数 h 的模型，并利用数据来估计这个模型中的未知参数。从而得到每个 h_i 的估计值，表示为 $\hat{h}_i$。在 GLS 变换中用 $\hat{h}_i$ 取代 h_i 就得到一个估计量，称为可行的 **GLS（FGLS）估计量**（feasible GLS estimator）。可行的 GLS 有时又被称为**估计的 GLS** 或 **EGLS**。 267

虽然有多种模型化异方差性的方法，但我们将学习一种特殊的、相当灵活的方法。假定

$$\mathrm{Var}(u \mid \mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k) \tag{8.30}$$

式中，x_1，x_2，…，x_k 为回归模型［参见方程（8.1）］中出现的自变量；δ_j 为未知参数。虽然也有可能出现 x_j 的其他函数形式，但我们将主要考虑式（8.30）。用上一小节的记号，即 $h(\mathbf{x}) = \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$。

你可能想知道我们为什么会使用式（8.30）中的指数函数。毕竟在用 Breusch-Pagan 检验来检验异方差性时，我们曾假定异方差性是 x_j 的一个线性函数。虽然像式（8.12）那样的线性形式在用于检验异方差性时没什么问题，但在用加权最小二乘法对异方差性进行修正时却会有问题。我们以前曾遇到过产生这个问题的原因：线性模型不能保证预测值都为正，而为了进行 WLS，要求我们估计的方差必须为正。

如果参数 δ_j 已知，那我们就只须像上一小节那样直接应用 WLS。可惜这种情况是不很现实的。最好是先用数据去估计这些参数，然后再使用这些参数的估计值来构造权数。如何估计 δ_j 呢？实质上，我们将借助一个小小的修改，把这个方程变换成一个可用 OLS 估计的线性形式。

在假定（8.30）之下，我们可以写

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)\ v$$

式中，v 以 $\mathbf{x} = x_1$，x_2，…，x_k 为条件时的均值等于 1。如果我们假定 v 确实与 $\mathbf{x}$ 无关，那我们就可以写成

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + e \tag{8.31}$$

式中，e 的均值等于零并与 x 无关；此方程中的截距与 δ_0 也不同，但无关紧要。因变量是平方误的对数。由于式（8.31）满足高斯-马尔科夫假定，所以我们可利用 OLS 得到 δ_j 的无偏估计量。

和平常一样，我们必须以 OLS 残差来取代观测不到的 u。所以，我们做回归

$$\log(\hat{u}^2)\text{ 对 } x_1, x_2, \cdots, x_k \tag{8.32}$$

我们实际上想从这个回归中得到的只是[$E(u^2)$或 $E(\hat{u}^2)$]拟合值：称之为 $\hat{g}_i$。然后，h_i 的估计值无非就是

$$\hat{h}_i = \exp(\hat{g}_i) \tag{8.33}$$

我们现在利用以 $1/\hat{h}_i$ 为权数的 WLS，并把步骤概括如下。

268

纠正异方差性的一个可行的 GLS 程序

1. 将 y 对 x_1，x_2，…，x_k 做回归并得到残差 $\hat{u}$。
2. 通过先将 OLS 残差进行平方，然后再取自然对数而得到 $\log(\hat{u}^2)$。
3. 做方程（8.32）中的回归并得到拟合值 $\hat{g}$。
4. 求出方程（8.32）中拟合值的指数：$\hat{h} = \exp(\hat{g})$。
5. 以 $1/\hat{h}_i$ 为权数用 WLS 来估计方程

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

如果能在 WLS 程序中使用 h_i 而不是 $\hat{h}_i$，那我们就知道，我们的估计量将是无偏的；事实上，假定我们已经正确地对异方差性建模，那它们还将是最优线性无偏估计量。不得不用同样的数据去估计 h_i 就意味着 FGLS 估计量不再是无偏估计量（因此也不可能是 BLUE）。不过，FGLS 估计量仍是一致的，而且比 OLS 更渐近有效。由于对方差参数的估计，所以很难证明这一点。但如果我们忽略这一点（确实可以忽略），其证明就类似于证明 OLS 是定理 5.3 中一类估计量中最有效的。无论如何，对于大样本的情况，当有迹象表明异方差性使 OLS 估计值的标准误变大时，FGLS 就成为 OLS 的一个颇具吸引力的替代方法。

我们必须记住，FGLS 估计量是方程

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

中参数的估计量。正如 OLS 估计值度量了每个 x_j 对 y 的边际影响一样，FGLS 的估计值也是如此。我们之所以用 FGLS 估计值取代 OLS 估计值，是因为前者更有效，而且其相关的检验统计量具有 t 和 F 分布，至少在大样本中如此。如果我们对方程（8.30）中所确定的方差有某种疑虑的话，可以对数据变换后的方程使用异方差稳健的标准误和检验统计量。

另一种估计 h_i 的有用方法是，用 OLS 拟合值及其平方取代回归式（8.32）中的自变量。换句话说，从

$$\log(\hat{u}^2)\text{ 对 } \hat{y}, \hat{y}^2 \tag{8.34}$$

的回归中得到拟合值 $\hat{g}_i$，然后与方程（8.33）中所做的完全一样求 $\hat{h}_i$。在前面的程序中，只有第（3）步发生了变化。

如果我们利用回归式（8.32）去估计方差函数，你可能想知道是否能仅用同样的这个回归去检验异方差性呢（可以使用 F 或 LM 检验）。实际上，帕克（Park，1966）曾提出这种思路。不幸的是，与8.3节所讨论过的检验相比，Park检验有一些问题。首先，其虚拟假设必须比同方差性更强，即 u 和 x 必须相互独立。Breusch-Pagan或White检验则没有这种要求。其次，利用OLS残差 $\hat{u}$ 取代式（8.32）中的 u，即便在大样本容量的情况下，也可
269 能导致 F 统计量偏离 F 分布。在我们已经讨论过的其他检验中则没有这个问题。出于这些原因，在检验异方差性时，我们不主张使用Park检验。回归式（8.32）能很好地为加权最小二乘效劳的原因在于，我们只需要 δ_j 的一致估计量，而回归式（8.32）当然能够做到。

例 8.7　对香烟的需求

我们利用SMOKE.RAW中的数据来估计一个对日香烟消费量的需求函数。由于大多数人不吸烟，所以因变量 *cigs* 对大多数观测都等于零。因为线性模型能导致负的预测值，所以不太理想。不过，通过一个线性模型，我们仍能对吸烟的决定因素有所了解。

用普通最小二乘法估计出来的方程（圆括号中给出了通常的OLS标准误）为

$$
\begin{aligned}
\widehat{cigs} = {} & -3.64 + 0.880\log(income) - 0.751\log(cigpric) \\
& (24.08)\ (0.728) \qquad\qquad\qquad (5.773) \\
& -0.501educ + 0.771age - 0.009\,0age^2 - 2.83restaurn \quad (8.35) \\
& (0.167) \qquad\quad (0.160) \qquad (0.001\,7) \qquad (1.11) \\
& n = 807, R^2 = 0.052\,6
\end{aligned}
$$

式中，*cigs* 为每天吸烟的数量；*income* 为年收入；*cigpric* 为每包香烟的价格（以美分为单位）；*educ* 为受教育年数；*age* 为（以年为单位的）年龄；*restaurn* 为一个二值变量（若此人居住的州禁止在餐馆吸烟，则取值1，否则取值0）。由于我们还要做加权最小二乘估计，所以就不再报告OLS的异方差稳健标准误。（顺便提一句，807个拟合值中有13个小于零；占样本中不到2%的比例，而且也不值得特别考虑。）

式（8.35）中的收入和香烟价格都不是统计显著的，而且它们的影响实际上也不大。比如，如果收入提高10%，预计 *cigs* 提高(0.880/100)(10)＝0.088，或者说，每天增加不到1/10根香烟。价格影响的幅度也与此类似。

每多受一年教育，就使平均每天吸烟的数量减少半根，而且这个影响是统计显著的。吸烟量还与年龄呈二次函数关系。在年龄达到 $age = 0.771/[2(0.009)] \approx 42.83$ 岁之前，吸烟量随年龄的增长而增加，随后则随年龄的增长而减少。这个二次函数的两项都是统计显著的。禁止在餐馆吸烟使吸烟量几乎每天平均减少3根。

270 方程 (8.35) 背后的误差项含有异方差性吗？将 OLS 残差的平方对式 (8.35) 中的自变量做 Breusch-Pagan 回归［参见方程 (8.14)］，得到 $R_{\hat{u}^2}^2=0.040$。虽然这么小的 R-平方看上去标志着没有异方差性，但我们必须记住计算 F 或 LM 统计量。如果样本容量很大，那么一个看起来很小的 $R_{\hat{u}^2}^2$ 也能导致对同方差性强有力的拒绝。LM 统计量为 $LM=807(0.040)=32.28$，而且这是一个 χ_6^2 随机变量的结果。p 值低于 0.000 015，这是异方差性的极强证据。

因此，我们利用前面的可行 GLS 程序再估计这个方程。所估计的方程为

$$
\begin{aligned}
\widehat{cigs} = {} & \underset{(17.80)}{5.64} + \underset{(0.44)}{1.30}\log(income)\ \ \underset{(4.46)}{2.94}\log(cigpric) - \underset{(0.120)}{0.463}\,educ \\
& + \underset{(0.097)}{0.482}\,age - \underset{(0.000\,9)}{0.005\,6}\,age^2 - \underset{(0.80)}{3.46}\,restaurn \qquad (8.36)\\
& n=807,\ R^2=0.113\,4
\end{aligned}
$$

现在收入效应既是统计显著的，而且在数量上也更大。价格效应也明显更大，但仍不是统计显著的。［其原因之一在于，*cigpric* 只随样本中不同的州而变化，所以 log(*cigpric*) 的变异性比 log(*income*)，*educ* 和 *age* 都要小得多。］

其他变量的估计值自然也多少有些变化，但基本情况仍然相同。吸烟量与学校教育负相关，与年龄呈二次关系，并且与餐馆禁止吸烟的限制负相关。

在用 WLS 估计之后，为了检验多重假设，就需要计算 F 统计量，这时我们必须小心从事。（不论是使用 F 统计量的残差平方形式还是 R-平方形式，都要小心。）重要之处在于，在估计不受约束模型和受约束模型时要采用相同的权数。我们应该首先用 OLS 估计不受约束模型。一旦得到了权数，就可以将它们用来估计受约束模型。F 统计量也就可以像平常那样计算。幸运的是，许多回归软件包在 WLS 估计之后，都有一个简单的命令来检验联合约束，所以我们不需要亲自做受约束模型的回归。

问题 8.4

假设方程 (8.30) 中异方差性的模型不正确，但我们还是基于这个方差使用了可行的 GLS 程序。WLS 仍是一致的，但通常的标准误、t 统计量等都不再成立，即使在渐近意义上也是如此。那我们又该怎么做呢？［提示：参见方程 (8.26)，如果 $\mathrm{Var}(u\ \ \boldsymbol{x})\neq\sigma h(\boldsymbol{x})$，$u_i^*$ 也包含了异方差性。］

例 8.7 暗示了一个在应用加权最小二乘时有时出现的问题：OLS 和 WLS 估计值可能相差甚远。在吸烟需求方程中，由于所有的系数都保持相同的符号，而且只有那些用 OLS 估计时在统计上不显著的变量才变化最大，所以这不是什么大问题。由于抽样误差的存在，所以 OLS 和 WLS 估计值总有所不同，问题是它们的差异是否足以改变重要的结论。

如果 OLS 和 WLS 得到符号不同而又都统计显著的估计值（比如，OLS 价

格弹性为正而又显著，而 WLS 价格弹性为负而又显著），或者估计值数量上的差异确实很大，那我们就应该表示怀疑。具体而言，这标志着其他高斯-马尔科夫假定之一是错误的，特别是对误差的零条件均值假定(MLR.3)。u 和任何一个自变量之间的相关都会导致 OLS 和 WLS 的偏误和不一致，而且偏误的大小通常都不同。**Hausman 检验**(Hausman，1978)可用来规范地比较 OLS 和 WLS 估计值，以看出其差距是否超过了抽样误差的范围。这个检验超出了本书的范围。在多数情况下，对估计值非正式地“打量”就足以发现问题。 271

8.5 再议线性概率模型

如我们在 7.6 节所见，当因变量是一个二值变量时，除非所有的斜率参数都是零，否则模型就一定包含异方差性。我们现在开始处理这个问题。

处理线性概率模型中异方差性问题的最简单方法，就是继续使用 OLS 估计，但也要计算检验统计量的稳健标准误。这就忽略了我们实际上知道 LPM 的异方差形式这个事实。不过，LPM 的 OLS 估计值很简单，而且常常能得到令人满意的结果。

例 8.8　已婚妇女的劳动力参与

在 7.6 节的劳动力参与的例子[见方程(7.29)]中，我们报告了通常的 OLS 标准误。现在我们同时计算异方差稳健的标准误，并报告在通常的标准误下面的方括号中：

$$
\begin{array}{lllll}
inlf = 0.586 & - 0.003\,4\,nwifeinc & + 0.038\,educ & + 0.039\,exper & \\
\quad (0.154) & (0.001\,4) & (0.007) & (0.006) & \\
\quad [0.151] & [0.001\,5] & [0.007] & [0.006] & \\
\quad - 0.000\,60\,exper^2 & - 0.016\,age & \quad 0.262\,kidslt6 & + 0.013\,0\,kidsge6 & \\
\quad (0.000\,18) & (0.002) & (0.034) & (0.013\,2) & \\
\quad [0.000\,19] & [0.002] & [0.032] & [0.013\,5] & \\
\quad n = 735,\ R^2 = 0.264 & & & & (8.37)
\end{array}
$$

有几个稳健标准误在报告的精度上和 OLS 标准误中都相同；所有的差异实际上都很小。因此，尽管异方差性在理论上是个问题，但在实践中并不是什么问题，至少对于本例而言如此。结果常常是，通常的 OLS 标准误和检验统计量与异方差的标准误和检验统计量相类似。此外，要计算二者还要花一定的时间。

一般而言，OLS 估计量在 LPM 中都是低效的。回忆 LPM 中 y 的条件方差为

272

$$\mathrm{Var}(y \mid \boldsymbol{x}) = p(\boldsymbol{x})[1 - p(\boldsymbol{x})] \tag{8.38}$$

其中

$$p(\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{8.39}$$

为响应概率（$y = 1$ 的成功概率）。概率 p（$\boldsymbol{x}$）显然取决于未知的总体参数 β_j。不过，我们确实得到了这些参数的无偏估计量，即 OLS 估计量。把 OLS 估计量代入方程（8.39），我们就得到 OLS 拟合值。因此，对每个观测 i，$\mathrm{Var}(y_i \mid \boldsymbol{x}_i)$ 由

$$\hat{h}_i = \hat{y}_i(1 - \hat{y}_i) \tag{8.40}$$

估计出来，其中 $\hat{y}_i$ 为第 i 个观测的 OLS 拟合值。现在，我们就像在 8.4 节那样应用可行的 GLS。

不幸的是，能对每个 i 估计 h_i 并不意味着我们就能直接进行 WLS 估计。原因就是我们在 7.6 节简要讨论的问题：拟合值 $\hat{y}_i$ 不一定落在单位区间内。如果 $\hat{y}_i < 0$ 或 $\hat{y}_i > 1$，方程（8.40）表明 $\hat{h}_i$ 为负。由于 WLS 程序对第 i 个观测总要乘以 $1/\sqrt{\hat{h}_i}$，所以如果某个观测的 $\hat{h}_i$ 为负（或等于零），那么这个方法就宣告失败。换句话说，WLS 的所有权数都必须为正。

在某些情况下，对于所有的 i，都有 $0 < \hat{y}_i < 1$，此时就可以用 WLS 估计 LPM。在那些观测数据很多而成功或失败的概率都很小的情形中，发现某些拟合值位于单位区间之外颇为常见。若是这样，像方程（8.37）中劳动力参与一例中所做的那样，最容易的办法就是放弃 WLS 并报告异方差稳健的统计量。另一种办法是调整那些小于 0 或大于 1 的拟合值，然后用于 WLS。一种建议是在 $\hat{y}_i < 0$ 时取 $\hat{y}_i = 0.01$，而在 $\hat{y}_i > 1$ 时取 $\hat{y}_i = 0.99$。不幸的是，这就要求研究者任意选择——比方说，为什么不用 0.001 和 0.999 作为调整值？如果多数拟合值都位于单位区间之外，对拟合值的调整就能影响结论；在这种情况下，仅仅使用 OLS 可能是最好的办法。

用加权最小二乘法估计线性概率模型

1. 用 OLS 估计模型并得到拟合值 $\hat{y}$。

2. 判断是否所有的拟合值都位于单位区间之内。如果是这样，就进行第（3）步。不然，则需要进行某种调整而使所有的拟合值都位于单位区间内。

3. 构造方程（8.40）中的估计方差。

4. 以 $1/\hat{h}$ 为权数用 WLS 估计方程

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

273 **例 8.9　学生拥有个人计算机的决定因素**

我们利用 GPA.RAW 中的数据，来估计学生拥有计算机的概率。用 *PC* 表示一个二值变量，若学生拥有一台计算机则取值 1，否则取值 0。变量 *hsGPA* 表示其高中时的 *GPA*，*ACT* 表示能力测验分数，而 *parcoll* 也是一个二值变量，若父母中至少有一人读过大学则取值 1。（将父母是否读过大学

分开考虑，不会得到个别显著的结论，因为它们完全是高度相关的。）

用OLS估计出来的方程是

$$
\begin{aligned}
\widehat{PC} = &-0.000\,4 + 0.065hsGPA + 0.000\,6ACT + 0.221parcoll \\
&(0.490\,5) \quad (0.137) \quad (0.015\,5) \quad (0.093) \\
&[0.488\,8] \quad [0.139] \quad [0.015\,8] \quad [0.087] \\
&n = 141, R^2 = 0.041\,5
\end{aligned}
\tag{8.41}
$$

恰如例8.8一样，通常的标准误和稳健标准误之间没有明显差异。不过，我们还用WLS估计了这个模型。因为所有的OLS拟合值都在单位区间之内，所以就无须进行调整：

$$
\begin{aligned}
\widehat{PC} = &0.026 + 0.033hsGPA + 0.004\,3ACT + 0.215parcoll \\
&(0.477) \quad (0.130) \quad (0.015\,5) \quad (0.086) \\
&n = 141,\ R^2 = 0.046\,4
\end{aligned}
\tag{8.42}
$$

OLS和WLS估计值之间没有重大差异。惟一显著的解释变量是*parcoll*，而且我们在两种情况下都估计出，如果父母中有一方读过大学，拥有计算机的概率约高出0.22。

▶ 小 结

我们首先回顾了普通最小二乘法在出现异方差时的性质。异方差性虽然不会导致OLS估计量的偏误和不一致性，但却导致通常的标准误和检验统计量都不再成立。我们说明了如何计算异方差稳健的标准误和t统计量，这些在多数回归软件包中都会例行计算。大多数回归软件包还会计算一个异方差稳健的F统计量。

我们讨论了检验异方差性的两种常见方法：Breusch-Pagan检验和White检验的一个特例。这两个统计量都涉及将OLS残差的平方对自变量（BP）或拟合值和拟合值的平方（White）做回归。如此则得到一个简单的F检验是渐近有效的；还有这些检验的拉格朗日乘数形式。 271

出现异方差性时，OLS就不再是最优线性无偏的估计量。若知道异方差性的形式，则可使用广义最小二乘（GLS）估计，这就使得加权最小二乘成为得到BLUE估计量的一种方法。WLS估计得到的检验统计量，或者在误差正态分布时确切有效，或者在误差非正态分布时渐近有效。当然，这里假定我们已经用了正确的异方差性模型。

更常见的是，我们在应用WLS之前必须先估计一个异方差模型。由此得到的可行的GLS估计量不再是无偏的，但它仍是一致和渐近有效的。WLS回归所得到的通常的统计量也是渐近有效的。我们又讨论了一种确保对所有观测的估计方差都严格为正的方法，这正是应用WLS的前提。

如我们在第 7 章所讨论的那样，二值因变量的线性概率模型必然具有异方差的误差项。处理这个问题的一个简单办法，就是计算异方差稳健的统计量。另外，如果所有的拟合值（即所估计的概率）都严格地介于 0～1 之间，那么就可以利用加权最小二乘法得到渐近有效的估计量。

关键术语

异方差性的 Breusch-Pagan 检验(BP 检验)
可行的 GLS(FGLS)估计量
广义最小二乘(GLS)估计量
未知形式的异方差
对异方差稳健的标准误
对异方差稳健的 F 统计量
对异方差稳健的 LM 统计量
对异方差稳健的 t 统计量
加权最小二乘(WLS)估计量
异方差性的 White 检验

习　题

8.1　下面哪种情况是异方差性造成的结果？

(i) OLS 估计量 $\hat{\beta}_j$ 是不一致的。

(ii) 通常的 F 检验不再服从 F 分布。

(iii) OLS 估计量不再是 BLUE。

8.2　考虑如下解释每月啤酒消费量的线性模型：

$$\begin{aligned}
&beer = \beta_0 + \beta_1 inc + \beta_2 price + \beta_3 educ + \beta_4 female + u \\
&\mathrm{E}(u \mid inc, price, educ, female) = 0 \\
&\mathrm{Var}(u \mid inc, price, educ, female) = \sigma^2 inc^2
\end{aligned}$$

写出将它变换成一个误差具有同方差性的方程。

8.3　判断正误：当模型中遗漏了重要变量时，WLS 优于 OLS。

8.4　利用 GPA3.RAW 中的数据，对秋季第二学期的学生估计了如下方程：

275

$$\begin{aligned}
\widehat{trmgpa} = {} & -2.12 + 0.900 crsgpa + 0.193 cumgpa + 0.0014 tothrs \\
& \quad (0.55) \quad (0.175) \qquad (0.064) \qquad (0.0012) \\
& \quad [0.55] \quad [0.166] \qquad [0.074] \qquad [0.0012] \\
& + 0.0018 sat - 0.0039 hsperc + 0.351 female - 0.157 season \\
& \quad (0.0002) \qquad (0.0018) \qquad (0.085) \qquad (0.098) \\
& \quad [0.0002] \qquad [0.0019] \qquad [0.079] \qquad [0.080] \\
& n = 269, \quad R^2 = 0.465
\end{aligned}$$

式中，$trmgpa$ 为本学期的 GPA；$crsgpa$ 为所修全部课程加权平均的 GPA；

tothrs 为此学期前总学分时；*sat* 表示 SAT 分数；*hsperc* 为其在高中班级中排名的百分位；*female* 为一个性别虚拟变量；*season* 也是一个虚拟变量，并在该学生在秋季有体育课时取值 1。通常的标准误和异方差稳健的标准误分别报告于圆括号和方括号中。

（i）变量 *crsgpa*，*cumgpa* 和 *tothrs* 都有预期的估计效应吗？这些变量中有哪些在 5% 的显著性水平上是统计显著的？使用不同的标准误，是否有什么影响？

（ii）为什么虚拟假设 H_0：$\beta_{crsgpa}=1$ 有意义？利用这两种标准误，在 5% 的显著性水平上针对双侧对立假设检验这个虚拟假设。陈述你的结论。

（iii）利用两种标准误来检验有体育课对学期 *GPA* 是否有影响。拒绝虚拟假设的显著性水平与所用的标准误有关系吗？

8.5　变量 *smokes* 是一个二值变量，如某人吸烟则取值 1，否则取值 0。利用 SMOKE.RAW 中的数据，我们估计了 *smokes* 的一个线性概率模型：

$$
\begin{array}{llll}
\widehat{smokes} = 0.656 - 0.069\log(cigpric) & + 0.012\log(income) & - 0.029educ & \\
\quad(0.855)\quad(0.204) & \quad(0.026) & \quad(0.006) & \\
\quad[0.856]\quad[0.207] & \quad[0.026] & \quad[0.006] & \\
+ 0.020age & - 0.000\,26age^2 & - 0.101restaurn & - 0.026white \\
\quad(0.006) & \quad(0.000\,06) & \quad(0.039) & \quad(0.052) \\
\quad[0.005] & \quad[0.000\,06] & \quad[0.038] & \quad[0.050] \\
n = 807, R^2 = 0.062 & & &
\end{array}
$$

如果被调查者是白人，则变量 *white* 等于 1，否则等于 0；其他的自变量都与例 8.7 中的定义一致。通常的标准误和异方差稳健的标准误都报告了。

（i）这两组标准误之间有什么重要差异吗？

（ii）保持其他因素不变，如果多受四年教育，预计其吸烟的概率会发生什么变化？

（iii）在哪个年龄以后，每大一岁都会降低吸烟的概率？

276　（iv）解释二值变量 *restaurn*（一个虚拟变量，如果此人所居住的州有饭馆禁止吸烟的限制则取值 1）的系数。

（v）数据集中第 206 个人具有如下特征：*cigpric* = 67.44，*income* = 6 500，*educ* = 16，*age* = 77，*restaurn* = 0，*white* = 0 和 *smokes* = 0。计算出此人吸烟的预计概率，并对结果进行评论。

计算机习题

8.6　利用 SLEEP75.RAW 中的数据估计如下睡眠方程：

$$
\begin{aligned}
sleep = {} & \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid \\
& + \beta_6 male + u
\end{aligned}
$$

（i）写出一个模型，容许 u 的方差在男人和女人之间有所不同。但这个方差不依赖于其他因素。

（ii）估计异方差模型中的参数。（你必须先用 OLS 估计 *sleep* 方程，以

得到 OLS 残差。）u 的估计方差对于男人或女人是否更高？

（iii）u 的方差是否对于男人和女人显著不同？

8.7 （i）利用 GPRICE1.RAW 中的数据得到方程（8.17）的异方差稳健标准误。讨论其与通常的标准误之间的任何重要差异。

（ii）对方程（8.18）重复第（i）步操作。

（iii）此例对异方差性和对因变量所做的变换说明了什么？

8.8 在方程（8.18）中应用异方差性的完全 White 检验［参见方程（8.19）］。利用 χ 平方形式的统计量并计算 p 值。你得到什么结论？

8.9 本题使用 VOTE1.RAW 中的数据。

（i）估计一个以 $voteA$ 为因变量并以 $prtystrA$，$democA$，$\log(expendA)$ 和 $\log(expendB)$ 为自变量的模型。得到 OLS 残差 $\hat{u}_i$，并将这些残差对所有的自变量进行回归。解释你为什么得到 $R^2=0$。

（ii）现在计算异方差性的 Breusch-Pagan 检验。使用 F 统计量的形式并报告 p 值。

（iii）同样利用 F 统计量形式计算异方差性的特殊 White 检验。现在异方差性的证据有多强？

8.10 本题使用 PNTSPRD.RAW 中的数据。

（i）变量 $sprdcvr$ 是一个二值变量，若拉斯维加斯队在大学篮球比赛中与其他队的分数差达到预定的差距，则此变量取值 1。$sprdcvr$ 的期望值（比方说 μ）表示在一场随机抽取的比赛中预定的分数差得以保持的概率。在 10% 的显著性水平上相对于 H_1：$\mu \neq 0.5$ 检验 H_0：$\mu = 0.5$，并讨论你的结果。［提示：将 $sprdcvr$ 只对一个截距项进行回归便得到一个 t 统计量，利用这个 t 统计量很容易做出来。］

（ii）553 个样本中有多少场比赛是在中立场地上进行的？

277 （iii）估计线性概率模型

$$sprdcvr = \beta_0 + \beta_1 favhome + \beta_2 neutral + \beta_3 fav25 + \beta_4 und25 + u$$

并以通常的形式报告结论。（报告通常的标准误和异方差稳健的标准误。）哪个变量在实际上和统计上都是最显著的？

（iv）解释为什么在虚拟假设 H_0：$\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ 下，模型中不存在异方差性。

（v）利用通常的 F 统计量检验第（iv）部分的虚拟假设，你得到什么结论？

（vi）给定上述分析，你会不会认为利用赛前可利用的信息系统地预测拉斯维加斯预定的分数差能否保得住是可能的？

8.11 在例 7.12 中，我们估计了一个线性概率模型，以说明一个年轻人在 1986 年是否被拘捕：

$$\begin{aligned} arr86 = {} & \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 tottime + \beta_4 ptime86 \\ & + \beta_5 qemp86 + u \end{aligned}$$

(i) 用OLS估计此模型，并验证其全部估计值都严格地介于0～1之间。最大和最小的估计值各是多少？

(ii) 像8.5节所讨论的那样，用加权最小二乘法估计这个方程。

(iii) 用WLS估计值决定*avgsen*和*tottime*在5%的显著性水平上是否联合显著。

8.12　本题利用LOANAPP.RAW中的数据。

(i) 估计习题7.16第（iii）部分中的方程，计算其异方差稳健的标准误。将β_{white}的95%的置信区间与非稳健的置信区间相比较。

(ii) 从第（i）部分的回归计算拟合值。其中有没有哪个估计值小于0？有没有哪个估计值大于1？这些情况对加权最小二乘估计的应用意味着什么？

第 9 章　模型设定和数据问题的深入探讨

278　我们在第 8 章讨论了高斯-马尔科夫假定不成立的一种情况。虽然误差的异方差性也可以被看成是一种模型误设，但它只是相对次要的一种。异方差性的出现并不会导致 OLS 估计量的偏误或不一致性，而且调整置信区间、t 统计量和 F 统计量，以便在 OLS 估计之后作出正确的推断，也相当容易；甚至通过使用加权最小二乘法还能得到更有效的估计量。

在本章，我们来讨论误差 u 和一个或多个解释变量相关这个严重得多的问题。记得在第 3 章，无论什么原因，如果 u 与 x_j 解释变量相关，那我们就称 x_j 为一个**内生解释变量**（endogenous explanatory variable）。我们还要进行更详尽的讨论，以说明一个解释变量为什么会内生的三个原因；在某些情形下，我们又探讨了可能的补救措施。

我们在第 3 章和第 5 章已经看到，遗漏一个关键变量能导致误差与某些解释变量之间的相关，从而通常导致所有的 OLS 估计量都是偏误和不一致性。在遗漏的变量是模型中的一个解释变量的函数的特殊情形下，模型就存在**函数形式误设**（functional form misspecification）的问题。

我们在第一节首先讨论函数形式误设所造成的后果，以及如何对它进行检验。在 9.2 节，我们说明代理变量的使用如何能解决或至少减轻遗漏变量的偏误。在 9.3 节，我们推导并解释在特定形式的**测量误差**（measurement error）下所引起的 OLS 偏误。9.4 节则讨论数据问题。

本章所有的估计程序都以 OLS 估计为基础。如我们将所见，导致误差与某些解释变量之间相关的某些问题，不可能通过对单一个横截面数据使用 OLS 而解决。我们把另外一些估计方法的处理推后到第 3 部分。

9.1 函数形式误设

如果一个多元回归模型没有正确地解释因变量和所观测到的解释变量之间的关系，那它就存在函数形式误设的问题。比如，若小时工资由 $\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 exper^2+u$ 决定，但遗漏了工作经历的平方项 $exper^2$，我们就会遇到一个函数形式误设的问题。我们从第 3 章已经知道，这通常会导致 β_0、β_1 和 β_2 有偏误的估计量。（因为 $exper^2$ 已从模型中剔除掉，所以不估计 β_3。）因此，错误地设定 $exper$ 是如何影响 279 $\log(wage)$ 的，这通常要导致教育回报 β_1 的估计量有偏误。偏误的大小取决于 β_3 的大小与 $educ$，$exper$ 和 $exper^2$ 之间的关系。

在估计工作经历的回报时情况更糟：即使能得到 β_2 的一个无偏估计量，我们也不可能估计出工作经历的回报，因为它等于 $\beta_2+2\beta_3 exper$（以小数形式）。仅使用 β_2 的无偏估计量可能起误导作用，特别是在 $exper$ 取极端值时。

作为另一个例子，假设 $\log(wage)$ 方程是

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 exper^2+\beta_4 female+\beta_5 female\cdot educ+u \tag{9.1}$$

式中，$female$ 为一个二值变量。如果遗漏了交互项 $female\cdot educ$，那么就误设了函数形式。一般而言，我们不可能得到其他任何一个参数的无偏估计量，而且由于教育回报取决于性别，所以遗漏交互项后，我们在估计什么的回报并不清楚。

遗漏自变量的函数并不是模型出现函数形式误设的惟一方式。例如，如果式（9.1）是满足前四个高斯-马尔科夫假定的真实模型，但我们却用 $wage$ 而不是 $\log(wage)$ 作为因变量，就不能得到偏效应的无偏或一致估计量。虽然以下检验对这种函数形式误设问题有一定的侦查能力，但还有更好的检验，我们在下一小节对非嵌套的不同函数形式进行检验时将会提到。

错误地设定一个模型的函数形式肯定会导致严重的后果。不过，从一个重要的方面来看，这个问题也是次要的：按定义，要得出一个很好地拟合数据的函数关系，我们已经有了所有必要变量的数据。这一点可与下一节提出的问题形成鲜明的对照，因为在下一节，将有一个我们不能搜集到其数据的关键变量被省略掉。

我们已经有了一个强有力的工具来侦查误设函数形式：联合排除性约束的 F 检验。通常，在模型中添加任何一个显著变量的平方项并进行一个联

合显著性检验，都是讲得通的。如果所增加的平方项是显著的，那就可以把它们放到模型中（代价是对模型的解释更复杂些）。然而，显著的平方项又可能是函数有其他形式这一问题的征兆，比如在该用变量的对数值时却使用了水平值，或反之。很难确定函数形式误设的准确原因。幸运的是，在许多情形下，使用某些变量的对数形式和添加二次项，就足以发现经济学中许多重要的非线性关系。

例 9.1 犯罪的经济模型

表 9.1 包含了犯罪经济模型（参见例 8.3）中的 OLS 估计值。我们首先估计了不含二次项的模型，其结果见于第（1）列。在第（2）列中，则增加了 *pcnv*，*ptime*86 和 *inc*86 的平方项；我们之所以选择这些变量的平方项，是因为它们在第（1）列中都是显著的。由于变量 *qemp*86 是一个只取五个值的离散变量，所以我们在第（2）列中也没有包括它的平方项。

问题 9.1 我们在表 9.1 的第（2）列中为什么没有包括 *black* 和 *hispan* 的平方项？

280

表 9.1 **因变量：*narr*86**

自变量	(1)	(2)
$pcnv$	−0.133 (0.040)	0.533 (0.154)
$pcnv^2$	—	0.730 (0.156)
$avgsen$	−0.011 (0.012)	−0.017 (0.012)
$tottime$	0.012 (0.009)	0.012 (0.009)
$ptime86$	−0.041 (0.009)	0.287 (0.004)
$ptime86^2$	—	−0.029 6 (0.003 9)
$qemp86$	−0.051 (0.014)	0.014 (0.017)
$inc86$	0.001 5 (0.000 3)	−0.003 4 (0.0008)
$inc86^2$	—	0.000 007 (0.000 003)
$black$	0.327 (0.045)	0.292 (0.045)
$hispan$	0.194 (0.040)	0.164 (0.039)
截距项	0.596 (0.036)	0.505 (0.037)
观测次数	2 725	2 725
R-平方	0.072 3	0.103 5

这些平方项中的每一个都是显著的，而且它们联合在一起是非常显著的（$df=3$ 和 2 713 时 $F=31.37$；p 值几乎为零）。所以，看起来原模型忽视了 281 某些潜在重要的非线性关系。

二次项的出现使得对模型的解释多少有些困难。比如，$pcnv$ 不再具有严格的阻止效应：$narr86$ 和 $pcnv$ 之间的关系在 $pcnv=0.365$ 之前为正，此后为负。我们可能会认为，在 $pcnv$ 的值较小时，只有很小甚至没有阻碍效应；在先前定罪率较高时才开始起作用。为了验证这个结论，我们必须使用比二次项更复杂的函数形式。可能 $pcnv$ 并非完全外生。比如，那些在过去未被定罪的人（故 $pcnv=0$）可能会偶尔犯罪，所以他们在 1986 年被拘捕的可能性也较小。这就使估计值产生偏误。

类似地，$narr86$ 和 $ptime86$ 之间的关系在 $ptime86=4.85$（几乎有两个月在坐牢）之前为正，此后为负。由于样本中的大多数人 1986 年都没有坐过牢，所以我们又必须小心解释这些结论。

合法收入在 $inc86=242.85$ 之前对 $narr86$ 都具有负效应；由于收入以百美元度量，所以这意味着年收入为 24 285 美元。样本中只有 46 个人的收入高于这个水平。于是，我们能断定 $narr86$ 和 $inc86$ 负相关，且具有递减的影响。

由于因变量的性质，例 9.1 是一个棘手的函数形式问题。还有一些其他模型，在理论上更适合于处理只取几个整数值的因变量。我们将在第 17 章简要地讨论这些模型。

对函数形式误设问题的一般检验：RESET

为了侦查一般的函数形式误设，已经提出了一些检验方法。事实表明拉姆齐（Ramsey，1969）的回归**设定误差检验**（regression specification error test，RESET）在这方面很有用。

RESET 背后的思想相当简单。如果原模型

$$y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+u \tag{9.2}$$

满足 MLR.3，那么在方程（9.2）中添加自变量的非线性关系应该是不显著的。在例 9.1 中，我们添加了显著解释变量的二次项。尽管这样做通常能侦查出函数形式误设，但如果原模型中有许多解释变量，它又有使用掉大量自由度的缺陷（很像异方差性的 White 检验的直接形式那样，要消耗大量的自由度）。此外，添加二次项还不能得到被忽略的某些特定非线性关系。而 RESET 则在方程（9.2）中添加 OLS 拟合值的多项式，以侦查函数形式误设的一般形式。

为了实施 RESET，我们必须决定在一个扩大回归中包括多少个拟合值

的函数。虽然对这个问题没有正确的回答，但在大多数应用研究中，都表明平方项和三次方项很有用。

令 $\hat{y}$ 表示估计式（9.2）所得到的 OLS 拟合值。考虑扩大方程

$$y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+\delta_1\hat{y}^2+\delta_2\hat{y}^3+\text{误差项} \tag{9.3}$$

这个方程看起来有些奇怪，因为原估计的拟合值的函数现在却作为解释变量出现。实际上，我们对式（9.3）中的估计参数并不感兴趣；我们只是利用这个方程来检验式（9.2）是否漏掉了重要的非线性关系。要记住，$\hat{y}^2$ 和 $\hat{y}^3$ 282 都只是 x_j 的非线性函数。

虚拟假设是，式（9.2）是正确设定的形式。于是，RESET 就是在扩大模型（9.3）中检验 H_0：$\delta_1=0$，$\delta_2=0$ 的 F 统计量。显著的 F 统计量则表明存在某种函数形式的问题。在大样本情况下，F 统计量的分布在虚拟假设（和高斯-马尔科夫假定）下渐近服从 $F_{2,n-k-3}$。扩大模型（9.3）的 df 为 $n-k-1-2=n-k-3$。也可以使用 LM 检验（χ 平方分布的 df 为 2），而且利用 8.2 节中讨论的方法，还可以进行对异方差稳健的检验。

例 9.2　住房价格方程

利用 HPRICE1.RAW 中的数据，我们估计了两个住房价格模型。第一个方程的所有变量都是水平值：

$$price=\beta_0+\beta_1 lotsize+\beta_2 sqrft+\beta_3 bdrms+u \tag{9.4}$$

第二个方程使用了除 $bdrms$ 外所有变量的对数形式：

$$lprice=\beta_0+\beta_1 llotsize+\beta_2 lsqrft+\beta_3 bdrms+u \tag{9.5}$$

利用 HPRICE3.RAW 中 $n=88$ 所住房，结果方程（9.4）的 RESET 统计量为 4.67；它就是随机变量 $F_{2,82}(n=88,k=3)$ 的值，相应的 p 值为 0.012。这就是式（9.4）中存在函数形式误设的证据。

式（9.5）的 RESET 统计量为 2.56，p 值为 0.084；于是，我们在 5% 的显著性水平上不能拒绝式（9.5）（尽管在 10% 的显著性水平上会拒绝）。基于 RESET，式（9.5）中的对数—对数模型更好。

283 在上例中，我们尝试了两个解释住房价格的模型。一个被 RESET 所拒绝，而另一个则未被拒绝（至少在 5% 的显著性水平上）。通常事情并没有这么简单。RESET 的一个缺陷是，当模型被拒绝后，它不能为我们该怎么做提供一个现实的方向。利用 RESET 拒绝式（9.4）并非立即表明式（9.5）就是下一步。之所以估计式（9.5），是因为常弹性模型易于解释，且具有良好的统计性质。在此例中，它还碰巧通过了函数形式检验。

有人声称 RESET 是模型误设的一个很一般性的检验，包括对观测不到的遗漏变量和异方差性的检验。不幸的是，如此使用 RESET 多属误导。可以证明，只要被遗漏的变量的期望值是模型中所包括自变量的线性函数，

RESET就无法侦查出变量遗漏问题［准确的表述可参见 Wooldridge（1995）］。而且，如果正确设定了函数形式，RESET对于侦查异方差性就无能为力。归根到底，RESET就是一个函数形式检验，别的什么都不是。

对非嵌套模型的检验

寻求对函数形式误设的其他类型（比如，试图决定某一自变量究竟应以水平值形式还是对数形式出现）作出检验，使我们离开了经典假设检验的辖域。有可能要相对模型

$$y=\beta_0+\beta_1\log(x_1)+\beta_2\log(x_2)+u \tag{9.6}$$

检验模型

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u \tag{9.7}$$

或者把这两个模型反过来［由于中英文表达习惯的不同，这里将原文中的式（9.6）和式（9.7）对调了——译者注］。然而，它们是非嵌套模型（参见第6章），所以我们不能仅使用标准的 F 检验。已经有人提出了两种不同的方法。一种方法是构造一个综合模型，将每个模型都作为一个特殊情形而包含其中，然后检验导致每个模型的约束。在目前的例子中，综合模型就是

$$y=\gamma_0+\gamma_1x_1+\gamma_2x_2+\gamma_3\log(x_1)+\gamma_4\log(x_2)+u \tag{9.8}$$

我们首先检验 H_0：$\gamma_3=0$，$\gamma_4=0$，作为对式（9.7）的检验。我们也可以检验 H_0：$\gamma_1=0$，$\gamma_2=0$，作为对式（9.6）的检验。这种方法由米曾和理查德（Mizon and Richard，1986）提出。

另一种方法由戴维森和麦金农（Davidson and MacKinnon，1981）提出。他们指出，如果式（9.7）是正确的，那么从另一个模型（9.6）得到的拟合值在式（9.7）中应该是不显著的。因此，为了检验式（9.7），我们首先用OLS估计模型（9.6）以得到拟合值，并记之为 $\check{y}$。然后，**Davidson-MacKinnon 检验**（Davidson-MacKinnon test）则基于方程

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\theta_1\check{y}+\text{误差项}$$

中 $\check{y}$ 的 t 统计量。（相对双侧对立假设）显著的 t 统计量则是对式（9.7）的拒绝。

284 类似地，如果 $\hat{y}$ 表示估计式（9.7）所得到的拟合值，那么对式（9.6）的检验，就是模型

$$y=\beta_0+\beta_1\log(x_1)+\beta_2\log(x_2)+\theta_1\hat{y}+\text{误差项}$$

显著的 t 统计量就是拒绝式（9.6）的证据。这同样的两个检验可用于检验任意两个具有相同因变量的非嵌套模型。

非嵌套检验也有一些问题。首先，不一定会出现一个明显好的模型。两

个模型可能都被拒绝，也可能没有一个被拒绝。在后一种情形中，我们可以使用调整 R-平方来进行选择。如果两个模型都被拒绝，则有更多的工作要做。不过，重要的是知道使用这种或那种形式的实际后果：如果关键自变量对 y 的影响没有多大差异，那么使用哪个模型实际上并不要紧。

第二个问题是，比方说用 Davidson-MacKinnon 检验拒绝了模型（9.6），这并不意味着式（9.7）就是正确的模型。模型（9.6）可能会因多种误设的函数形式而被拒绝。

一个更为可能的问题是，在因变量不同的模型进行比较时，如何得到非嵌套检验。典型的情况就是，一个因变量是 y，一个因变量是$\log(y)$。我们在第 6 章看到，用刚刚得到的拟合优度度量来进行比较，需要小心从事。虽然有人提出了解决这个问题的检验，但对其分析超出了本书的范围。[有一种易于理解并便于实施的检验，可参见 Wooldridge（1994a）。]

9.2 对观测不到的解释变量使用代理变量

当一个模型通常因缺乏数据而排除了一个关键变量时，就会出现更困难的问题。考虑一个明确提出能力（*abil*）会影响$\log(wage)$的工资方程：

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 exper+\beta_3 abil+u \tag{9.9}$$

这个模型明确表示了我们在度量 *educ* 和 *exper* 的回报时想保持能力不变。如果 *educ* 与 *abil* 相关，那么将 *abil* 放到误差项中去就会导致 β_1（和 β_2）的 OLS 估计量有偏误，这已经是一个反复出现的问题了。

我们在方程（9.9）中主要关心的是斜率参数 β_1 和 β_2。我们实际上并不关心是否能得到截距 β_0 的无偏或一致估计量；如我们稍后将看到的那样，这通常是不可能得到的。此外，由于观测不到 *abil*，所以我们从未想过去估计 β_3；事实上，由于能力充其量只是一个模糊的概念，所以我们也不知道如何去解释 β_3。

我们怎样才能解决（或至少减小）像式（9.9）这样的方程中因遗漏变量而导致的偏误呢？一种可能是找到遗漏变量的一个**代理变量**（proxy variable）。大致说来，代理变量就是某种与我们在分析中试图控制而又观测不到的变量相关的东西。在工资方程中，一种可能性就是用智商或 *IQ* 作为能力变量的一个工具。这并不要求 *IQ* 就等同于能力，只需要 *IQ* 与能力相关，
285 我们在下面的分析中加以说明。

全部的主要思想都可以用一个有三个自变量的模型加以说明，其中有两个自变量是可以观测的：

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3^*+u \tag{9.10}$$

我们假定有 y，x_1 和 x_2 的数据 [在工资方程的例子中，分别是$\log(wage)$， *educ* 和 *exper*]。虽然解释变量 x_3^* 观测不到，但我们有 x_3^* 的一个代理变量，并称之为 x_3。

我们对 x_3 有什么要求呢？起码它应该与 x_3^* 有某种关系。这一点由简单回归方程

$$x_3^* = \delta_0 + \delta_3 x_3 + v_3 \tag{9.11}$$

表示。式中，v_3 为因 x_3^* 与 x_3 并非完全相关所导致的误差。参数 δ_3 度量了 x_3^* 与 x_3 之间的关系；特别是，我们认为 x_3^* 和 x_3 正相关，所以 $\delta_3>0$。如果 $\delta_3=0$，则 x_3 不是 x_3^* 合适的代理变量。式（9.11）中的截距 δ_0，无非是容许 x_3^* 和 x_3 以不同的尺度来度量。（比方说，在美国这个总体中，肯定不能要求观测不到的能力变量与 *IQ* 具有相等的平均值。）

我们如何能够利用 x_3 得到 β_1 和 β_2 的无偏（或至少是一致）估计量呢？建议假装认为 x_3 就是 x_3^*，所以我们做

$$y \text{ 对 } x_1,\ x_2,\ x_3 \text{ 的回归} \tag{9.12}$$

因为我们在做 OLS 之前，只是用 x_3 取代了 x_3^*，所以我们称之为**遗漏变量问题的植入解**（plug-in solution to the omitted variables problem）。如果 x_3 与 x_3^* 真正相关，看起来是一件不错的事情。不过由于 x_3 与 x_3^* 并不相同，所以我们应该决定这个做法什么时候确实能给出 β_1 和 β_2 的一致估计量。

植入解能得到 β_1 和 β_2 的一致估计量所需要的假定可区分为对 u 的假定和对 v_3 的假定：

（1）误差 u 与 x_1，x_2 和 x_3^* 都不相关，这就是对模型（9.10）的一个标准假定。而且，u 与 x_3 也不相关。后一个假定只是意味着，一旦总体模型中包括了 x_1，x_2 和 x_3^*，x_3 就无足轻重了。既然 x_3 是 x_3^* 的一个代理变量，所以根据定义，实质上的确如此：是 x_3^* 而不是 x_3 影响 y。因此，假定 u 与 x_1，x_2，x_3^* 和 x_3 都不相关不是很有争议。（对这个假定的另一种表述是，给定所有这些变量，u 的期望值为零。）

（2）误差 v_3 与 x_1，x_2 和 x_3 都不相关。v_3 与 x_1，x_2 不相关的假定要求 x_3 是 x_3^* 的一个"好"代理变量。最容易看出这一点的办法是，用条件期望的术语类似地写出这些假定：

$$\mathrm{E}(x_3^* \mid x_1, x_2, x_3) = \mathrm{E}(x_3^* \mid x_3) = \delta_0 + \delta_3 x_3 \tag{9.13}$$

286 最重要的第一个等式表明，一旦控制了 x_3，x_3^* 的期望值就与 x_1 或 x_2 无关。换句话说，一旦排除了 x_3 的影响，x_3^* 就与 x_1 和 x_2 零相关。

在工资方程（9.9）中，其中 *IQ* 是能力的代理变量，条件（9.13）就变成

$$\mathrm{E}(abil \mid educ, exper, IQ) = \mathrm{E}(abil \mid IQ) = \delta_0 + \delta_3 IQ$$

因此，能力的平均水平只随 *IQ* 而变化，而不随 *educ* 和 *exper* 而变化。这种说法合理吗？或许不完全正确，但与真实情况可能很接近。在工资方程中包

括 IQ 以看教育的估计回报有什么变化，当然值得一做。

我们很容易看出，前面的假定足以使植入解发挥作用。如果我们将方程（9.11）代入方程（9.10）并做简单运算，则得到

$$y = (\beta_0 + \beta_3\delta_0) + \beta_1 x_1 + \beta_2 x_2 + \beta_3\delta_3 x_3 + u + \beta_3 v_3$$

将这个方程中的合成误差记为 $e = u + \beta_3 v_3$；它取决于我们所关心的模型（9.10）中的误差和代理变量方程中的误差 v_3。由于 u 和 v_3 的均值都为零且都与 x_1，x_2，x_3 无关，所以 e 的均值也为零并与 x_1，x_2，x_3 无关。将这个方程写成

$$y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e$$

式中，$\alpha_0 = (\beta_0 + \beta_3\delta_0)$ 为新的截距项；$\alpha_3 = \beta_3\delta_3$ 为代理变量 x_3 的斜率参数。像我们前面提到的那样，当我们做（9.12）中的回归时，不会得到 β_0 和 β_3 的无偏估计量；但我们却能得到 α_0，β_1，β_2 和 α_3 的无偏（或至少是一致）估计量。重要的是，我们能得到参数 β_1 和 β_2 很好的估计值。

在多数情形中，α_3 的估计值实际上比 β_3 的估计值更有意义。比如，在工资方程中，度量了在 IQ 的分数多一分时的工资回报。由于多数总体的 IQ 分布都现成可用，所以有可能知道 IQ 在其他条件不变情况下对工资的影响。

例 9.3　*IQ* 作为能力的代理变量

布莱克本和努马克（Blackburn and Neumark，1992）中的文件 WAGE2.RAW，包含了 935 名男人在 1980 年的月薪、受教育程度、几个人口统计方面的变量和 IQ 值等信息。作为解决能力遗漏偏误的一种方法，我们在一个标准对数工资方程中增加 IQ 变量。结果在表 9.2 中给出。

我们感兴趣的是，在增加 IQ 变量后，教育的估计回报会有什么变化。第（1）列包含了不用 IQ 作为代理变量时的估计值。此时教育的估计回报为 6.5%。如果我们认为被省略的能力变量与 *educ* 正相关，那么我们就会认为这个估计值偏高。（更准确地说，所有随机样本的平均值将过高。）当在方程中增加 IQ 后，教育回报下降到 5.4%，这印证了我们先前对遗漏能力变量之偏误的信念。

287 最近，在赫姆斯坦和默里（Herrnstein and Murray，1994）合著的一本颇具争议的书《贝尔曲线（Bell curve)》中，详细引证了 IQ 对社会经济产出的影响。第（2）列表明，在控制了其他几个因素后，IQ 对工资所得确实具有统计显著的正影响。在其他条件相同的情况下，IQ 提高 10 分，预计会使月工资增加 3.6%。因为在美国总体中 IQ 的标准差为 15，所以 IQ 提高一个标准差，就会伴随着收入 5.4% 的提升。这就等同于多受一年教育对工资的预期影响。从第（2）列显然可见，教育在提高工资方面仍具有重要作用，尽管其影响没有原估计那么大。

在第（1）列和第（2）列中还有其他一些有趣的观测。在方程中增加

IQ 只使 R-平方从 0.253 提高到 0.263。log($wage$)中大部分的变异都没有被第 (2) 列中的因素所解释。此外，在方程中增加 IQ 也未能消除黑人和白人男性在估计工资上的差异：一个在 IQ、受教育水平、工作经历等方面都和白人一样的黑人，预计其工资所挣比白人约少 14.3%，而且这个差异是相当统计显著的。

表 9.2 **因变量：log($wage$)**

自变量	(1)	(2)	(3)
$educ$	0.065	0.054	0.018
	(0.006)	(0.007)	(0.041)
$exper$	0.014	0.014	0.014
	(0.003)	(0.003)	(0.003)
$tenure$	0.012	0.011	0.011
	(0.002)	(0.002)	(0.002)
$married$	0.199	0.200	0.201
	(0.039)	(0.039)	(0.039)
$south$	0.091	−0.080	−0.080
	(0.026)	(0.026)	(0.026)
$urban$	0.184	0.182	0.184
	(0.027)	(0.027)	(0.027)
$black$	−0.188	−0.143	−0.147
	(0.038)	(0.039)	(0.040)
IQ	—	0.003 6	−0.000 9
		(0.001 0)	(0.005 2)
$educ \cdot IQ$	—	—	0.000 34
			(0.000 38)
截距	5.395	5.176	5.648
	(0.113)	(0.128)	(0.546)
观测次数	935	935	935
R-平方	0.253	0.263	0.263

288 表 9.2 中的第 (3) 列还包括了交互项 $educ \cdot IQ$，这就使得 $educ$ 和 $abil$ 在决定log($wage$)的过程中有相互影响的可能性。我们可能会认为，能力越强的人，其教育回报就越高，但事实证明并非如此：交互项不显著，且交互项的引入在使模型复杂化的同时，还使得 $educ$ 和 IQ 个别地看也不显著。因此，第 (2) 列的估计值更为可取。

问题 9.2

就表 9.2 第 (3) 列中 $educ$ 的系数较小且在统计上不显著，你能得出什么结论？[提示：当方程中出现 $educ \cdot IQ$ 时，对 $educ$ 的系数该做何解释？]

在本例中没有理由只使用能力的一个代理变量即算完事。数据集 WAGE2.RAW 还包含了每个人在工作领域内知识（knowledge of the world of work，KWW）测验中的分数。这就给出对能力的一种不同度量，它可以取代 *IQ* 或与 *IQ* 一起用于估计教育回报（参见习题 9.7）。

容易看出，如果代理变量不符合前述假定，使用一个代理变量为何仍将导致偏误。假使与式（9.11）不同，观测不到的变量 x_3^* 与所有可观测变量都相关：

$$x_3^* = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + v_3 \tag{9.14}$$

式中，v_3 的均值为零且与 x_1，x_2，x_3 都不相关。方程（9.11）假定 δ_1 和 δ_2 都为零。通过将方程（9.14）代入方程（9.10），我们得到

$$y = (\beta_0 + \beta_3\delta_0) + (\beta_1 + \beta_3\delta_1)x_1 + (\beta_2 + \beta_3\delta_2)x_2 + \beta_3\delta_3 x_3 + u + \beta_3 v_3 \tag{9.15}$$

由此得到$\text{plim}(\hat{\beta}_1) = \beta_1 + \beta_3\delta_1$ 和 $\text{plim}(\hat{\beta}_2) = \beta_2 + \beta_3\delta_2$。［得到这个结果的原因是，式（9.15）中的误差 $u + \beta_3 v_3$ 的均值为零，且与 x_1，x_2 和 x_3 都不相关。］在前面 $x_1 = educ$ 和 $x_3^* = abil$ 的例子中 $\beta_3 > 0$，所以若 *abil* 和 *educ* 有正的偏相关（$\delta_1 > 0$），则 $\hat{\beta}_1$ 存在正的偏误。因此，如果 *IQ* 不是一个很好的代理变量，那么用 *IQ* 作为代理变量仍将得到教育回报的向上偏误。但我们有理由希望，这个偏误比我们完全忽略遗漏能力变量时小。

289 代理变量也可以以二值信息的形式出现。在例 7.9［参见方程（7.15）］中，我们讨论了克鲁格（Krueger，1993）对工作中使用计算机的回报的估计。克鲁格还包括了一个二值变量，以表示一个工人在家是否使用计算机（以及在工作中和在家使用计算机的一个交互项）。他在方程中把在家使用计算机情况包括进来的主要原因是作为观测不到的“技术能力”的代理变量，这种技术能力既能直接影响工资，又与工作中的计算机使用情况相关。

用滞后因变量作为代理变量

在某些应用研究中，如前面的工资一例，我们对要控制哪些观测不到的因素至少都会有些模糊的认识，这就使我们能够选择代理变量。但在另外一些应用研究中，我们猜测一个或多个自变量与遗漏变量相关，可是对如何得到遗漏变量的代理变量却一筹莫展。在这种情况下，我们可以将较早期的因变量值包括进来加以控制。这种方法对政策分析特别有用。

虽然在横截面方程中使用一个滞后因变量提高了对数据的要求，但也为解释导致因变量现期差异的历史因素找到了一个简单方法，而这种现期差异用其他方法都很难解释。比如，某些城市过去都曾有过高犯罪率。同时导致现在和过去犯罪率很高的观测不到的因素中，许多都是相同的。类似地，某些大学传统上都比其他大学的学术成就更胜一筹。惯性影响也是通过引入 y 的滞后值来刻画的。

考虑一个解释城市犯罪率的简单方程：

$$crime = \beta_0 + \beta_1 unem + \beta_2 expend + \beta_3 crime_{-1} + u \tag{9.16}$$

式中，$crime$ 为对人均犯罪次数的某种度量；$unem$ 表示城市失业率；$expend$ 为在执法方面的人均支出；$crime_{-1}$为以前某个年度的犯罪率（可以是上年或几年前的某一年）。我们感兴趣的是失业率和执法支出对犯罪率的影响。

在方程中包括 $crime_{-1}$的目的何在呢？当然，由于犯罪具有惯性，所以我们预期 $\beta_3>0$。但把它放到方程中的主要原因在于，历史上犯罪率高的城市对犯罪的预防也会花更多的钱。因此，我们（计量经济学家）观测不到而又影响 $crime$ 的因素，可能会与 $expend$（和 $unem$）相关。如果用一个纯粹的横截面分析，那我们不可能得到执法支出对犯罪率的因果效应的无偏估计量。但通过在方程中包括 $crime_{-1}$，我们至少可做如下试验：如果两个城市以前有相同的犯罪率，现在又有相同的失业率，那么 β_2 就度量了多增加执法支出 1 美元对犯罪率的影响。

例 9.4 城市犯罪率

我们来估计方程（9.16）中犯罪模型的一个常弹性形式（由于 $unem$ 是个百分数，所以保留其水平值形式）。CRIME2.RAW 中的数据来自 1987 年的 46 个城市。也有 1982 年的犯罪率数据可用，而我们在试图控制那些影响犯罪率而又可能与当年执法支出相关的不可观测因素时，将其用做另外一个自变量。表 9.3 包含了有关结果。 290

在方程中没有滞后犯罪率时，失业率和执法支出的影响都是违背直觉的：尽管$\log(lawexpc_{87})$的 t 统计量为 1.17，但没有一个是统计显著的。一种可能性是，增加的执法支出改善了报案手续，所以报告了更多的犯罪。但也有可能，近来犯罪率高的城市在执法支出上花了更多的钱。

表 9.3 **因变量：$\log(crmrte_{87})$**

自变量	(1)	(2)
$unem_{87}$	-0.029 (0.032)	0.009 (0.020)
$\log(lawexpc_{87})$	0.203 (0.173)	-0.140 (0.109)
$\log(crmrte_{82})$	—	1.194 (0.132)
截距	3.34 (1.25)	0.076 (0.821)
观测次数	46	46
R-平方	0.057	0.680

把5年前犯罪率的对数加进来，对支出系数具有很大的影响。犯罪率对支出的弹性变成 -0.14，且 $t=-1.28$。虽然这算不上十分显著，但它表明，一旦样本中有了更多的城市，一个更复杂的模型能够给出显著的结论。

无足为奇，现在的犯罪率与过去的犯罪率强相关。估计值表明，如果1982年的犯罪率高1%，那么预计1987年的犯罪率约高1.19%。我们不能拒绝现在的犯罪率对过去犯罪率的弹性一致的虚拟假设 $[t=(1.194-1)/0.132\approx1.47]$。虽然增加过去的犯罪率明显提高了回归的解释能力，但这没什么好奇怪的。把滞后犯罪率包括进来的基本原因，是为了在其他条件不变情况下估计 $\log(lawexpc_{87})$ 对 $\log(crmrte_{87})$ 的影响时，得到一个更好的估计值。

作为控制观测不到的变量的一般方法，将滞后的 y 放到方程中的做法，谈不上完美无缺。但在估计政策变量对发生的各种结果的影响时，它可用于得到一个更好的估计值。

291 添加 y 的滞后值并不是使用两年数据来控制遗漏变量的惟一办法。当我们在第13章和第14章讨论综列数据方法时，将探讨使用相同横截面单元在不同时点出现的重复数据的其他方法。

9.3 有测量误差的OLS性质

在经济应用研究中，我们有时候不能搜集到那些确实影响经济行为的变量的数据。一个很好的例子是，一个家庭在试图决定某年向慈善机构捐款多少时所面临的边际所得税率。很难得到边际税率，或者说，很难把各种收入水平的税率概括成一个数字。但我们可基于总收入和总税收支付计算出一个平均税率。

当我们在一个回归模型中使用经济变量不精确的度量时，我们的模型就包含了测量误差。在本节中，我们推导测量误差对普通最小二乘估计所造成的后果。虽然OLS在一些特定的假定之下是无偏的，但在其他情况下它又是有偏误的。在某些偏误的情形下，我们能推导出渐近偏误的大小。

如我们将看到的那样，测量误差问题与上一节讨论的遗漏变量—代理变量问题具有类似的统计结构，但它们在概念上是不同的。在代理变量情形中，我们在寻找一个与观测不到的变量多少有些联系的变量。在测量误差情形中，我们没能观测的变量却具有完好定义的定量含义（如边际税率或年收入），但我们对它测量的记录可能包含了误差。比如，报告的年收入是对实际年收入的一种测量，而 *IQ* 值却是能力的一个代理变量。

代理变量与测量误差问题之间另一个重要的差别在于：在测量误差问题中，被误测的自变量通常是主要的焦点之一；而在代理变量情形中，被遗漏变量的偏效应很少成为关注的核心，我们通常关心的是其他自变量的影响。

在我们详加考虑之前，我们应该记住，只有计量经济学家所能为之搜集数据的变量，与影响个人、家庭、企业等决策的变量不同时，测量误差才会成为问题。

因变量中的测量误差

我们首先从只有因变量存在测量误差的情况开始。令 y^* 表示我们（一如既往，在总体中）欲加以解释的变量，比如，y^* 可以是家庭年度储蓄。回归模型具有通常的形式

$$y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{9.17}$$

292 而且，我们假定它满足高斯-马尔科夫假定。我们令 y 表示所观测到的对 y^* 的度量。在储蓄一例中，y 就是所报告的年度储蓄。不幸的是，家庭在报告他们的家庭年度储蓄时并不完美，容易漏掉某些项目或高估对基金的贡献量。一般而言，我们可以想像 y 和 y^* 有所不同，至少对于家庭总体中的一部分来说如此。

（总体中的）测量误差被定义为观测值和实际值之差：

$$e_0 = y - y^* \tag{9.18}$$

对于总体中随机抽取的某个 i，我们可以写出 $e_{i0} = y_i - y_i^*$，但重要的一点是，总体中的测量误差是怎样与其他因素相关的。为了得到一个可估计的模型，我们把它改写成 $y^* = y - e_0$，并代入方程（9.17），整理得到

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u + e_0 \tag{9.19}$$

方程(9.19)中的误差项为 $u + e_0$。由于 $y, x_1, x_2, \cdots, x_k$ 都可以观测到，所以我们就能用 OLS 来估计这个模型。实质上，我们无非忽略了 y 是 y^* 的一个不完美度量这一事实，并像通常一样进行操作。

以 y 取代 y^* 在什么情况下能得到 β_j 的一致估计量呢？由于原模型(9.17)满足高斯-马尔科夫假定，所以 u 的均值为零且与每个 x_j 都不相关。惟一自然的做法就是假定测量误差的均值也为零；若不然，我们则得到截距 β_0 的一个有偏估计量，这很少成为值得忧虑的原因。更为重要的是，我们就测量误差 e_0 和解释变量 x_j 之间关系所做的假定。通常的假定是，y 的测量误差在统计上独立于每个解释变量。若果真如此，则从式(9.19)得到的 OLS 估计量就是无偏和一致的。进一步讲，通常的 OLS 推断程序（t、F 和 LM 统计量）也都生效。

若 e_0 和 u 像通常假定的那样不相关，则有 $\mathrm{Var}(u + e_0) = \sigma_u^2 + \sigma_0^2 > \sigma_u^2$。这意味着，因变量的测量误差导致误差方差比没有测量误差时更大；这当然也导致 OLS 估计量的方差更大。预计会是这样，但我们却无能为力（除非搜集更多的数据）。基本要点是，若测量误差与自变量不相关，那 OLS 估计就有良好的性质。

例 9.5　有测量误差的储蓄方程

考虑一个储蓄方程

$$sav^* = \beta_0 + \beta_1 inc + \beta_2 size + \beta_3 educ + \beta_4 age + u$$

但实际储蓄（sav^*）可能与报告的储蓄（sav）不一致。问题在于，sav 中测量误差的大小是否与其他变量系统相关。假定测量误差与 *inc*，*size*，*educ* 和 *age* 不相关可能是合理的。另一方面，我们也会想到，收入越高或受教育越多的家庭可能报告其储蓄时也越准确。除非我们能搜集到 sav^* 的更多数据，否则我们永远也不会知道测量误差是否和 *inc* 或 *educ* 相关；然后每个观测的测量误差都可以由 $e_{i0} = sav_i - sav_i^*$ 计算。 293

当因变量是对数形式时，即因变量为$\log(y^*)$时，测量误差方程自然就有如下形式：

$$\log(y) = \log(y^*) + e_0 \tag{9.20}$$

这种形式来自 y 的倍乘测量误差：$y = y^* a_0$，其中 $a_0 > 0$，$e_0 = \log(a_0)$。

例 9.6　废弃率中的测量误差

在 7.6 节讨论的一个例子中，我们想决定在职培训津贴是否会减少制造业企业里的废弃率。我们当然希望企业报告的废弃率在度量时不存在误差。（实际上，样本中的大多数企业甚至根本就不报告废弃率。）在一个简单的回归模型中，便可表述为

$$\log(scrap^*) = \beta_0 + \beta_1 grant + u$$

式中，$scrap^*$ 为真正的废弃率；*grant* 为该企业是否得到津贴的虚拟变量。测量误差方程为

$$\log(scrap) = \log(scrap^*) + e_0$$

测量误差 e_0 与企业是否得到津贴无关吗？犬儒主义者可能认为，得到津贴的企业为了说明津贴的有效性而更可能低报其废弃率。若是如此，则在可估计方程

$$\log(scrap) = \beta_0 + \beta_1 grant + u + e_0$$

中，误差 $u + e_0$ 就与 *grant* 负相关。这就导致 β_1 存在向下的偏误，从而倾向于使培训项目看起来比实际上更有效。（记住，由于提高的工人生产力与较低的废弃率相联系，所以 β_1 负得越多，就意味着项目越有效。）

本小节的起码结论是，如果因变量的测量误差与一个或多个解释变量系统相关，那就会导致 OLS 的偏误。如果测量误差像通常那样只是一个与解

释变量无关的随机报告误差，那么 OLS 就完全适用。

解释变量中的测量误差

294 从传统来看，某个解释变量中的测量误差被认为是比因变量中的测量误差严重得多的问题。在本小节，我们将看出为什么会是这样。

我们首先从简单的回归模型

$$y=\beta_0+\beta_1x_1^*+u \qquad (9.21)$$

开始，而且我们假定它至少满足前四个高斯-马尔科夫假定。这意味着，对式（9.21）进行 OLS 估计将得到 β_0 和 β_1 的无偏和一致估计量。问题是，x_1^* 观测不到。但我们有对 x_1^* 的一个测量值 x_1。比方说，x_1^* 是实际收入，x_1 而是报告收入。

总体中的测量误差无非是

$$e_1=x_1-x_1^* \qquad (9.22)$$

而其取值可能为正、负或零。假定总体中平均的测量误差为零：$E(e_1)=0$。做此假定很自然，而且无论如何也不会影响以下的重要结论。下面的分析将保持的假定是：u 与 x_1^* 和 x_1 都不相关。利用条件期望的形式，我们可以写成$E(y|x_1^*,x_1)=E(y|x_1^*)$，这个式子只是表明，在控制了 x_1^* 之后，x 不影响 y。我们使用了在代理变量情形中同样的假定，而且没有什么争议；根据定义，几乎总能成立。

我们想知道，如果我们仅以 x_1 取代 x_1^* 并将 y 对x_1 回归，那么 OLS 具有什么样的性质。其性质关键取决于我们对测量误差所做的假定。有两个假定已成为计量经济学文献中的焦点，而且它们都代表了截然不同的极端。第一个假定是，e_1 与所观测到的测量值 x_1 不相关：

$$\mathrm{Cov}(x_1,e_1)=0 \qquad (9.23)$$

根据式（9.22）中的关系，如果假定（9.23）正确，则 e_1 一定与观测不到的变量 x_1^* 相关。为了决定 OLS 在此情形下的性质，我们写出$x_1^*=x_1-e_1$ 并代入方程（9.21）：

$$y=\beta_0+\beta_1x_1+(u-\beta_1e_1) \qquad (9.24)$$

由于我们已经假定 u 和e_1 的均值都为零，且与 x_1 无关，所以 $u-\beta_1e_1$ 的均值也为零并与 x_1 无关。于是，以 x_1 取代 x_1^* 的 OLS 估计得到了 β_1（以及 β_0）的一致估计量。由于 u 与 e_1 无关，所以式（9.23）中误差的方差为 $\mathrm{Var}(u-\beta_1e_1)=\sigma_u^2+\beta_1^2\sigma_{e_1}^2$。因此，除非 $\beta_1=0$，否则测量误差就会提高误差方差。但这并不会影响任何一个 OLS 性质（除了 $\hat{\beta}_j$ 的方差比我们能直接观测到 x_1^* 时更大。）

295 e_1 与 x_1 无关的假定类似于我们在 9.2 节所做的代理变量假定。由于这个假定意味着 OLS 具有其全部优良性质，所以这并不是计量经济学家在提到解释变量的测量误差时所经常考虑的。**经典的含误差变量**（classical errors-in-variables，CEV）假定是指测量误差与观测不到的解释变量无关：

$$\mathrm{Cov}(x_1^*,e_1)=0 \tag{9.25}$$

将所观测到的测量值写成真正的解释变量与测量误差之和：

$$x_1=x_1^*+e_1$$

然后再假定 x_1 的两个成分无关，就得到了经典的含误差变量假定。（这个假定与对 u 的假定无关；我们仍保留 u 与 x_1^* 和 x_1 并因此与 e_1 无关的假定。）

如果假定（9.25）成立，那 x_1 与 e_1 就一定相关：

$$\mathrm{Cov}(x_1,e_1)=\mathrm{E}(x_1e_1)=\mathrm{E}(x_1^*e_1)+\mathrm{E}(e_1^2)=0+\sigma_{e_1}^2=\sigma_{e_1}^2 \tag{9.26}$$

因此，在 CEV 假定之下，x_1 与 e_1 之间的协方差就等于测量误差的方差。

根据方程（9.24），我们可以看出，x_1 与 e_1 相关将引起问题。由于 u 和 x_1 无关，所以 x_1 与合成误差 $u-\beta_1e_1$ 之间的协方差就是

$$\mathrm{Cov}(x_1,u-\beta_1e_1)=-\beta_1\mathrm{Cov}(x_1,e_1)=-\beta_1\sigma_e^2$$

因此，在 CEV 情形下，y 对 x_1 的 OLS 回归将给出一个无偏而又不一致的估计量。

利用第 5 章中的渐近结论，我们可以决定 OLS 中不一致的数量。$\hat{\beta}_1$ 的概率极限就是 β_1 加上 x_1 与 $u-\beta_1e_1$ 的协方差和 x_1 的方差之比：

$$\begin{aligned}\mathrm{plim}(\hat{\beta}_1)&=\beta_1+\frac{\mathrm{Cov}(x_1,u-\beta_1e_1)}{\mathrm{Var}(x_1)}\\&=\beta_1-\frac{\beta_1\sigma_{e_1}^2}{\sigma_{x_1}^{2}{}^*+\sigma_{e_1}^2}=\beta_1\Big(1-\frac{\sigma_{e_1}^2}{\sigma_{x_1}^{2}{}^*+\sigma_{e_1}^2}\Big)\\&=\beta_1\Big(\frac{\sigma_{x_1}^{2}{}^*}{\sigma_{x_1}^{2}{}^*+\sigma_{e_1}^2}\Big)\end{aligned} \tag{9.27}$$

式中我们用到了 $\mathrm{Var}(x_1)=\mathrm{Var}(x_1^*)+\mathrm{Var}(e_1)$。

方程（9.27）很有意思。用以乘 β_1 的项 $\mathrm{Var}(x_1^*)/\mathrm{Var}(x_1)$ 总小于 1［CEV 假定（9.25）的一个含义］。因此，$\mathrm{plim}(\hat{\beta}_1)$ 总比 β_1 更接近于零。这种情况被称为 OLS 因经典的含误差变量而导致的**衰减偏误**（attenuation bias）：平均而言（或在大样本中），所估计的 OLS 影响将会变小。特别是，
296 若 β_1 为正，则 $\hat{\beta}_1$ 倾向于低估 β_1。虽然这是一个重要结论，但它有赖于 CEV 结构。

如果 x_1^* 的方差相对于测量误差的方差很大，那么 OLS 中的不一致性将会很小。这是因为，当 $\sigma_{x_1}^{2}{}^*/\sigma_{e_1}^2$ 很大时，$\mathrm{Var}(x_1^*)/\mathrm{Var}(x_1)$ 将接近于 1。因此，取决于 x_1^* 的变异相对于 e_1 的变异而言究竟有多大，测量误差不一定导致很大的偏误。

当我们引入更多的解释变量时，情况就更复杂。为便于说明，考虑模型

$$y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + \beta_3 x_3 + u \tag{9.28}$$

其中三个解释变量中的第一个有测量误差。我们作出一个很自然的假定：u 与 x_1^*，x_2，x_3 和 x_1 都不相关。同样，关键的假定是针对测量误差 e_1 的假定。几乎在所有情况下，都假定 e_1 与不存在测量误差的解释变量 x_2 和 x_3 无关。关键问题是，e_1 是否与 x_1 无关。若无关，则 y 对 x_1，x_2 和 x_3 的 OLS 回归将得到一致估计量。这一点很容易看出来，只须将方程写成

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u - \beta_1 e_1 \tag{9.29}$$

式中，u 和 e_1 都与所有的解释变量无关。

在 CEV 假定（9.25）之下，因为在方程（9.29）中 e_1 和 x_1 相关，所以 OLS 将是有偏和不一致的。记住，这意味着，一般而言，所有的 OLS 估计量都是有偏的，而不仅仅是 $\hat{\beta}_1$。方程（9.27）中推导的衰减偏误会怎么样呢？结果在估计 β_1 时仍存在衰减偏误。可以证明

$$\operatorname{plim}(\hat{\beta}_1) = \beta_1\left(\frac{\sigma^2_{r_1^*}}{\sigma^2_{r_1^*} + \sigma^2_{e_1}}\right) \tag{9.30}$$

式中，r_1^* 表示方程 $x_1^* = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3 + r_1^*$ 中的总体偏误。当 x_1 是惟一被错误度量的变量时，公式（9.30）对 k 个变量的一般情形也成立。

对于估计不存在测量误差的变量的 β_j 来说，就没有那么明确的结论。在 x_1^* 与 x_2 和 x_3 都不相关的特殊情形中，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 都是一致的。但在实践中几乎不存在这种情况。一般说来，单一个变量的测量误差会导致所有估计量的不一致性。不幸的是，偏误的大小甚至方向都不容易得到。

例 9.7　存在测量误差的 *GPA* 方程

考虑估计家庭收入在控制了 *hsGPA* 和 *SAT* 后对大学 *GPA* 的影响这个问题。可能的情况是，尽管家庭收入在大学之前的表现很重要，但它对学生在大学里的表明没有直接影响。为检验这一点，我们可能会提出模型 297

$$colGPA = \beta_0 + \beta_1 faminc^* + \beta_2 hsGPA + \beta_3 SAT + u$$

式中，$faminc^*$ 为实际家庭年收入。（它可能会以对数形式出现，但为便于说明，我们采用水平值形式。）*colGPA*，*hsGPA* 和 *SAT* 的精确数据相对容易得到。但家庭收入，特别是学生报告的家庭收入，很容易被误测。如果 $faminc = faminc^* + e_1$，而且 CEV 假定成立，那么，用报告的家庭收入取代实际家庭收入，将使 β_1 的 OLS 估计量产生向 0 的偏误。其后果之一就是，对 H_0：$\beta_1 = 0$ 的检验中发现 $\beta_1 > 0$ 的可能性下降。

当然，测量误差也可能出现在不止一个解释变量中，或出现在某些解释变量和因变量之中。如我们前面所讨论的那样，通常假定因变量的测量误差

与所有的解释变量都无关，无论这些解释变量是否能观测到。在对 CEV 假定推广之后的条件下，推导 OLS 估计量的偏误很复杂，而且也不会得到明确的结论。

在某些情况下，式（9.25）中的 CEV 假定不可能正确。考虑例 9.7 的一个变形：

$$colGPA = \beta_0 + \beta_1 smoked^* + \beta_2 hsGPA + \beta_3 SAT + u$$

式中，*smoked** 为一名学生在过去 30 天内实际吸食大麻的次数；变量 *smoked* 表示学生对如下问题的回答：你在过去的 30 天内有多少次吸食过大麻？假设我们认为标准的测量误差模型是正确的：

$$smoked = smoked^* + e_1$$

即便我们假定学生都试图如实地回答，CEV 假定也不可能成立。完全不吸食大麻的学生（所以 $smoked^* = 0$）可能会报告 $smoked = 0$，所以对那些从不吸食大麻的学生而言，测量误差为零。当 $smoked^* > 0$ 时，学生极有可能会错算其在过去 30 天内吸食大麻的次数。这就说明，测量误差 e_1 和实际吸食大麻的次数 *smoked** 相关，这就违背了式（9.25）中的 CEV 假定。不幸的是，推导不满足式（9.23）或式（9.25）的测量误差的含义相当困难，而且超出了本书的研究范围。

问题 9.3 ☞

令 *educ** 表示以年度量的实际受教育程度（也可以是非整数），而令 *educ* 表示报告完成的最高学历。根据经典的含误差变量模型，你认为 *educ* 和 *educ** 相关吗？

在结束本节之前，我们强调一下，CEV 假定（9.25）先验地看，并不比意味着 OLS 估计量的一致性的假定（9.23）更好或更坏。事实可能介于二者之间，如果 e_1 同时与 x_1^* 和 x_1 相关，那么 OLS 就是不一致的。这就提出一个重要问题：在有经典的含误差变量或其他某种与 x_1 相关的测量误差的情况下，我们就一定得到不一致的估计量吗？幸运的是，回答是否定的。第 15 章表明，在某些假定下，即便出现了一般性的测量误差，也能一致地
298 估计出参数。我们把这方面的讨论推后，因为它要求我们离开 OLS 估计的辖域。

9.4 数据缺失、非随机样本和异常观测

上一节讨论的测量误差问题也可被看成一个数据问题：我们不能得到我们所关心的变量的数据。而且，在经典的含误差变量模型中，合成误差项与被误测的自变量相关，因而违背了高斯-马尔科夫假定。

我们在前几章反复讨论的另一个数据问题是解释变量之间的多重共线

性。记住，解释变量之间的相关并不违背任何假定。当两个自变量高度相关时，很难估计每个变量的偏效应。但这一点正确地反映在 OLS 统计量中。

在本节，我们对可能违背随机抽样假定 MLR.2 的数据问题加以介绍。我们可以分离出非随机抽样对 OLS 没有影响的情形。但在其他情形中，非随机抽样则会导致 OLS 估计量有偏和不一致。在第 17 章对这个问题进行了更完整的探讨，并证明了这里所做出的几个论断。

数据缺失

数据缺失（missing data）问题可能会以多种形式出现。我们通常对一群人、一些学校、一些城市等搜集一个随机样本，而以后又发现样本中有几个单位的某些关键变量的信息找不到。比如，在数据集 BWGHT.RAW 中，1 388个观测中有 197 个没有母亲、父亲或父母双方受教育的信息。在法学院起薪中位数的数据集 LAWSCH85.RAW 中，156 个学院中有 6 个没有报告其新生 *LSAT* 成绩中位数的信息；某些法学院还缺失其他变量的信息。

如果一个观测缺失了其因变量或一个自变量的数据，那么这个观测就不能用在多元回归分析中。实际上，如果正确地表明了缺失的数据，那么所有现代回归软件包都会跟踪缺失的数据，并在回归计算时简单地把相应观测忽略掉便是。我们在有关婴儿出生体重的例 4.9 中清楚地看到了这一点，由于缺少父母受教育方面的信息，有 197 个观测被去掉了。

除了减小回归可用的样本容量之外，数据缺失还有其他统计影响吗？那要取
299 决于数据缺失的原因。如果数据是随机缺失的，那么无非就是来自总体的可用的随机样本的容量减小了。尽管这使得估计量没那么准确，但也不会引人任何偏误：随机抽样假定 MLR.2 仍成立。虽然也有办法使用只缺失部分变量数据的观测的信息，但实践中通常不这样做。因为这样做对估计量的改进通常都很小，可使用的方法多少都有些复杂。在大多数情况下，我们就不用有信息缺失的观测。

非随机样本

当数据缺失导致样本变成总体的一个非随机样本时，就更成问题。比如，在婴儿出生体重的数据集中，若那些受教育程度低于平均水平的人缺失数据的概率更大的话，情况会怎么样？或者我们在 9.2 节使用了一个包括 *IQ* 值的工资数据集，而且在构造这个数据集时，去掉了样本中几个没有 *IQ* 值的人。如果高智商的人更容易得到其 *IQ* 值，那么这个样本就不能代表总体，也就违背了随机抽样假定 *MLR*.2，我们肯定会担心其对 OLS 估计的影响。

非随机抽样的某些特定类型也并不会导致 OLS 的偏误和不一致性。在

（没有 MLR.2 的）高斯-马尔科夫假定下，事实表明，样本可在自变量的基础上加以选择，而且不会导致任何统计问题。这就是**基于自变量的样本选择**，它是**外生样本选择**（exogenous sample selection）的一个例子。为便于说明，假设我们在估计一个储蓄函数，其中年度储蓄取决于收入、年龄、家庭规模及其他某些因素。一个简单的模型是

$$saving = \beta_0 + \beta_1 income + \beta_2 age + \beta_3 size + u \tag{9.31}$$

假设数据集是基于对 35 岁以上人群的调查，所以我们得到所有成年人的一个非随机样本。虽然不是很理想，但我们仍能利用这个非随机样本得到总体模型（9.31）中参数的无偏和一致估计量。我们在此不进行规范的证明，基于非随机样本的 OLS 估计无偏的原因在于，回归函数$\mathrm{E}(saving \mid income, age, size)$对由 *income*，*age* 或 *size* 所刻画的总体中的任何一个子集都是一样的。如果自变量在这个子总体中有充分的变化，那么基于这个自变量的选择除了导致低效估计外，就算不上什么严重问题。

在刚刚提到的 *IQ* 一例中，由于没有一个基于 *IQ* 的固定规则来确定某人是否属于样本，所以情况就没有那么明确。不过，进入样本的概率随着 *IQ* 的提高而提高。如果决定样本选择的其他因素独立于工资方程中的误差项，那么我们就得到另一种外生样本选择的情况，而使用选择样本的 OLS 在其他高斯-马尔科夫假定下也将具有其全部理想性质。

当选择基于因变量 y 时，情况就大不一样，这种情况被称为**基于因变量的样本选择**，也是**内生样本选择**（endogenous sample selection）的一个例子。如果基于因变量的值高于或低于某给定值而选择样本的话，OLS 在估计总体模型时就总会产生偏误。比如，假设我们想估计所有成人总体中个人财富与其他几个因素之间的关系：

300

$$wealth = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 age + u \tag{9.32}$$

假设只有财富不到 7.5 万美元的人才被包括在样本中。这是我们所关心总体的一个非随机样本，而且它是基于因变量的值而选择的。利用财富不足 7.5 万美元的人构成的样本将导致式（9.32）中参数的估计量有偏和不一致。简单而言，原因在于，总体回归$\mathrm{E}(wealth \mid educ, exper, age)$与以财富不足 7.5 万美元者为条件的期望均值并不相等。

其他的样本选择问题就更加微妙。例如，在以前几个例子中，我们曾估计了各种变量特别是受教育水平和工作经历对小时工资的影响。我们从头到尾所用的数据集 WAGE1.RAW 实质上是在工作的个人的一个随机样本。劳动经济学家通常对估计受教育水平对工资出价（wage offer）的影响感兴趣。其思想是，每个达到工作年龄的人都面临着一个小时工资报价，他或她既可以在那个工资下去工作，又可以不工作。对于那些工作的人来说，工资报价刚好就是其目前所挣的工资。对于那些不在工作的人来说，我们通常观察不到其工资报价。现在，由于工资报价方程

$$\log(wage^o) = \beta_0 + \beta_1 educ + \beta_2 exper + u \tag{9.33}$$

代表了所有达到工作年龄者构成的总体，所以我们就不可能用这个总体的一个随机样本来估计它；相反，我们只有那些正在工作的人的工资报价数据（尽管也有那些不在工作的人在 *educ* 和 *exper* 方面的数据）。如果我们用工薪者的一个随机样本去估计式（9.33），那我们能得到无偏的估计量吗？这种情况并不清楚。由于样本是基于个人的工作决策而选择的（与基于工资报价的高低相反），所以这与前面那种情形不同。不过，由于工作决策可能与那些影响工资报价的观测不到的因素相关，所以选择可能是内生的，而且可能导致 OLS 估计量中存在样本选择偏误。我们在第 17 章将讨论一些用于检验和纠正样本选择偏误的方法。

问题 9.4

假设我们对在任总统的竞选支出对选民支持率的影响感兴趣。某些在任总统作出不追求连任的选择。如果我们只能搜集到那些确实希望连任的在任者在投票支持率和支出方面的数据，那么这是有内生样本选择的可能吗？

异常观测

在某些应用研究中，特别是（但不仅仅是）在样本集较小时，OLS 估计值会受到一个或几个观测的影响。这种观测就被称为**异常数据**（outliers）或**有重要影响的观测**（influential observations）。粗略地讲，如果将一个观测从回归分析中去掉使得 OLS 估计值发生“很大”的实际变化，那么它就是一个异常数据。

OLS 对旁路观测敏感是因为它最小化了残差平方和：在最小二乘的最小化问题中，越大的（正或负）残差，其权数就越大。如果略微变动一下我们的样本，估计值就发生很大的变化，那么我们就要注意。

301 在统计学家和计量经济学家从理论上研究异常数据问题时，有时认为数据是来自给定总体（尽管其分布不同寻常，以致出现一些极端值）的一个随机样本，有时又假定异常数据来自一个不同的总体。从实践的观点来看，出现异常数据的原因有二。最易于处理的情况是，在输入数据时出了差错。在数字后面多加一个零或点错了小数点的位置，都会使 OLS 估计值出错，特别是在样本容量很小的时候。为了发现数据输入过程中的错误，计算摘要统计量（特别是最小值或最大值）总是一个好主意。不幸的是，数据的错误输入并不总是那么明显。

在从一个很小的总体中抽样时，如果总体中的一个或几个元素在某个重要方面与总体中其他元素差别很大，那也可能出现数据异常的问题。要决定在回归分析中留下还是去掉这种观测可能很困难，而且由此得到的估计量的统计性质也很复杂。异常观测也能通过提高解释变量的变异（减小标准误）而提供重要信息。但在一个或几个数据点会显著改变结论时，也许应该分别在包括和不包括这些异常观测的情况下各报告一个 OLS 结果。

例 9.8　R&D的强度与企业规模

假设R&D支出占销售额的百分比（*rdintens*）与企业销售额*sales*（以百万计）和利润在销售额的百分比（*profmarg*）相关：

$$rdintens = \beta_0 + \beta_1 sales + \beta_2 profmarg + u \tag{9.34}$$

利用RDCHEM.RAW中32家化工企业的数据估计的OLS方程为

$$\widehat{rdintens} = 2.625 + 0.000\,053 sales + 0.044\,6 profmarg$$
$$(0.586)(0.000\,044) \qquad (0.046\,2)$$
$$n = 32, R^2 = 0.076\,1, \bar{R}^2 = 0.012\,4$$

在此回归中，即便在10%的显著性水平上，*sales*和*profmarg*在统计上都不显著。

32家企业中，有31家的年销售额低于200亿美元，而有一家企业的年销售额接近400亿美元。图9.1表明了这家企业与样本中其他企业有多大的差距。从销售额看，这家企业的销售额超过了其他每家企业销售额的2倍以上，所以在不包括此点的情况下估计这个模型可能是个好主意。当我们这么做时，则得到

$$\widehat{rdintens} = 2.297 + 0.000\,186 sales + 0.047\,8 profmarg$$
$$(0.592)(0.000\,084) \qquad (0.044\,5)$$
$$n = 31, R^2 = 0.172\,8, \bar{R}^2 = 0.113\,7$$

如果将这家最大的企业从回归中去掉，那么*sales*的系数就超过原来系数的3倍，而且其*t*统计量现在也超过了2。利用较小企业的样本，我们将断定R&D与企业规模之间有统计显著的正效应。利润率仍不显著，而且其系数变化不大。

302

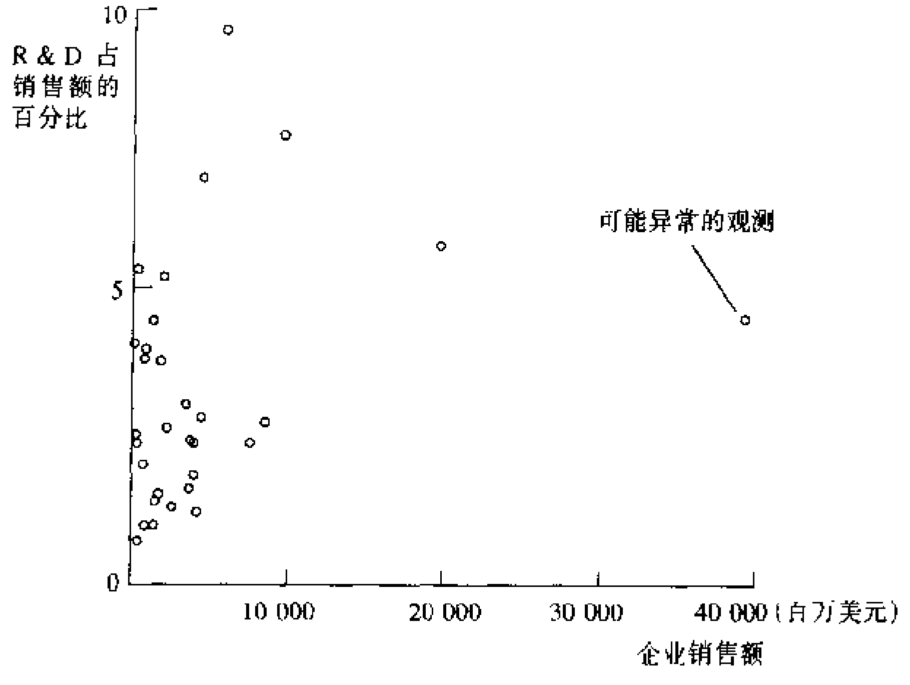


图9.1　R&D的强度相对企业规模的散点图

有时也用使用了所有观测的 OLS 回归中的残差大小来定义异常数据，但这不是一个好办法。在上例中，在回归中使用所有企业的话，有一个销售额略低于 46 亿美元的企业具有最大的残差（约 6.37）。最大企业的残差为 -1.62，与 0 之间的差别还不到估计的一个标准差（$\hat{\sigma}=1.82$）。而去掉这个残差最大的企业，根本就不会使结论有很大的改变。

某些函数形式对异常观测就没那么敏感。我们在 6.2 节中提到，对大多数经济变量来说，取对数会显著地缩小数据的取值范围，并得到能解释数据的取值范围更宽的函数形式（如常弹性模型）。

例 9.9 R&D 强度

我们可以用如下模型来检验 R&D 的强度是否随着企业规模的扩大而提高：

$$rd = sales^{\beta_1}\cdot\exp(\beta_0+\beta_2 profmarg+u) \tag{9.35}$$

303 于是，保持其他因素不变，当且仅当 $\beta_1>1$ 时，R&D 的强度才会随着销售额的增加而提高。对式（9.35）取对数得到

$$\log(rd)=\beta_0+\beta_1\log(sales)+\beta_2 profmarg+u \tag{9.36}$$

当我们使用所有 32 家企业时，回归方程是

$$\begin{aligned}\widehat{\log}(rd)&=-4.378+1.084\log(sales)+0.021\,7\,profmarg\\&\qquad(0.468)\quad(0.062)\qquad\qquad(0.012\,8)\\&\qquad n=32,R^2=0.918\,0,\bar{R}^2=0.912\,3\end{aligned}$$

而去掉最大的企业后得到的方程是

$$\begin{aligned}\widehat{\log}(rd)&=-4.404+1.088\log(sales)+0.021\,8\,profmarg\\&\qquad(0.511)\quad(0.067)\qquad\qquad(0.013\,0)\\&\qquad n=31,R^2=0.903\,7,R^2=0.896\,8\end{aligned}$$

这两个结论在实践中是一样的。我们不能在任何一种情形下相对于 $H_1:\beta_1>1$ 而拒绝虚拟假设 $H_0:\beta_1=1$。（为什么？）

在某些情形下，一开始就怀疑某特定观测与样本中所有其他观测根本不同。这种情况通常在我们使用诸如城市、县或州层次的加总数据时会发生。如下即是一例。

例 9.10 州婴儿死亡率

在《美国统计摘要》中可以得到婴儿死亡率、人均收入和保健（某种度量）的州一级数据。我们在此给出一个相当简单的分析，以说明异常数据的影响。我们得到美国 50 个州和哥伦比亚特区（D.C.）1990 年的数据。变量 *infmort* 表示每 1 000 个新生婴儿在出生后第一年的死亡人数，*pcinc* 表示人

均收入，*physic* 表示每 100 000 市民拥有的医生人数，而 *popul* 表示人口总数（以千人计）。我们将自变量都以其对数形式包含在模型中：

$$\begin{aligned}\widehat{infmort} = {} & 33.86 - 4.68\log(pcinc) + 4.15\log(physic)\\ & (20.43)(2.60) \qquad\qquad (1.51)\\ & \quad 0.088\log(popul)\\ & (0.287)\\ & n = 51, R^2 = 0.139, \bar{R}^2 = 0.084\end{aligned} \tag{9.37}$$

恰如所料，预计较高的人均收入会降低婴儿死亡率，但人均更多的医生数却伴随着更高的婴儿死亡率，这多少有些违背直觉。婴儿死亡率看上去与人口规模没什么关系。

304 由于哥伦比亚特区在一些小区域内极度贫穷，而在另一些小区域内有巨富，所以它与其他的州不同。实际上，1990 年 D.C. 的婴儿死亡率为 20.7，与之相对比，婴儿死亡率第二高的州也只有 12.4。D.C. 每 100 000 市民拥有的医生人数为 615，而第二高的州只有 337。哥伦比亚特区极高的医生比例和极高的婴儿死亡率肯定会影响估计结果。如果我们从回归中去掉 D.C.，则得到

$$\begin{aligned}\widehat{infmort} = {} & 23.95 - 0.57\log(pcinc) \quad 2.74\log(physic)\\ & (12.42)(1.64) \qquad\qquad (1.19)\\ & +0.629\log(popul)\\ & (0.191)\\ & n = 50, R^2 = 0.273, \bar{R}^2 = 0.226\end{aligned} \tag{9.38}$$

我们现在发现，人均更多的医生会降低婴儿死亡率，而且估计值在 5% 的显著性水平上统计显著地异于零。人均收入的影响急剧下降，而且不再具有统计显著的影响。在方程（9.38）中，在人口越多的州，婴儿死亡率越高，而且这种关系在统计上极为显著。此外，当从回归中去掉 D.C. 后，*infmort* 的变异被解释得更多。显然，D.C. 对原始估计值有相当明显的影响，我们在任何深入一步的分析中都可能要将它去掉。

不同于这种必须由个人来主观地决定特定观测的影响，那些能侦查出有影响观测的统计量有时也很有用。这些统计量确实存在，但超出了本书的研究范围。[比如，可参见 Belsley、Kuh and Welsch（1980）。]

在结束本节之前，我们提一下处理有影响观测的另一种方法。我们不再试图在应用最小二乘法之前就发现异常数据，而是可以使用一个比 OLS 对异常数据更不敏感的估计方法。这种方法之一就是**最小绝对离差**或 LAD。LAD 估计量最小化残差的绝对值之和，而不再是最小化残差平方和。与 OLS 相比，LAD 给很大的残差的权数较小。因此，少数观测的变化的影响就没有那么大。

尽管 LAD 有助于抑制异常数据的影响，但它也有某些缺陷。首先，没

有一个估计量的表达公式；它们只能通过计算机上的迭代方法才能求出。虽然对于今天强大的个人计算机而言不是很困难，但很大的数据集仍须进行耗时巨大的计算。其次，只有在误差项 u 对称分布时，LAD 才能一致地估计总体回归函数中的参数（在条件期望的意义上）。第三，如果误差项 u 是正态分布的，LAD 比 OLS 更缺乏（渐近）效率。当然，如果误差 u 真是正态分布的，那么得到一个很大的异常数据的概率就很小，我们也就可能对 OLS 表示满意

最小绝对离差正是通常所谓**稳健回归**的一个特殊情形。用统计术语来说，稳健的回归估计量对极端值相对缺乏敏感性：从本质上讲，对较大残差赋予的权数比最小二乘法中的权数要小。尽管这个性质是准确的，在这里使用“稳健”一语也可能导致混淆。如前面所提，为了在条件期望的意义上一致地估计参数，LAD 估计量要求误差项关于零对称分布。而 OLS 则不要求
305 这些。（回想一下，高斯-马尔科夫假定也没有包括误差项的对称分布要求。）

LAD 在条件中位数的意义上确实能一致地估计参数，而不管误差项是否对称分布。在某些情况下，这很有意思，但我们现在不再追究这一思想。伯克（Berk，1990）对稳健回归方法给出了入门性的介绍。

▶ 小结

我们又进一步研究了实证横截面分析中经常出现的一些重要的设定和数据问题。误设的函数形式使所估计的方程难以解释。不过，通过增加二次项、计算 RESET 或相对一个非嵌套的对立模型而使用 Davidson-MacKinnon 检验，就可以侦查出不正确的函数形式，而无须额外的数据搜集。

遗漏变量的解决就更困难一些。在 9.2 节，我们根据用代理变量去代替遗漏变量讨论了一种可能的解决办法。在一些合理的假定之下，在一个 OLS 回归中包括一个代理变量，会消除（或至少是减小）偏误。使用这种方法的困难在于难以找到代理变量，一般可能的办法是使用因变量前一年的数据。

应用经济学家常常很关心测量误差。在经典的变量误差（CEV）假定下，因变量的测量误差对 OLS 的统计性质没有什么影响。相反，在一个自变量的 CEV 假定下，被误测变量系数的 OLS 估计量会产生向零的偏误，而其他变量系数的偏误则可大可小，而且难以决定。

从一个基本总体的非随机抽样，可能会导致 OLS 的偏误。当样本选择与误差项 u 相关时，OLS 通常有偏且不一致。另一方面，外生样本选择（或者基于解释变量，或者独立于 u）并不会导致 OLS 出现问题。数据集中的异常数据可能对 OLS 估计值具有很大的影响，特别是在样本很小的情况下。这就要求至少是不规范地辨别出异常数据，并在排除被猜疑是异常数据的情况下重新估计模型。

关键术语

衰减偏误
经典的含误差变量（CEV）
戴维森-麦金农检验
内生解释变量
内生样本选择
外生样本选择
函数形式误设
有重要影响的观测值
滞后因变量

测量误差
数据缺失
倍乘测量误差
非嵌套模型
非随机样本
异常值
遗漏变量问题的植入解
代理变量
回归设定误差检验（RESET）

习　题

306 9.1　在习题 4.11 中，利用 CEOSAL2.RAW 中的数据估计模型

$$\log(salary) = \beta_0 + \beta_1\log(sales) + \beta_2\log(mktval) + \beta_3 profmarg + \beta_4 ceoten + \beta_5 comten + u$$

所得到的 $R^2 = 0.353$（$n = 177$）。若添加 $ceoten^2$ 和 $comten^2$ 后，$R^2 = 0.375$。此模型中是否有函数形式误设的证据？

9.2　让我们对习题 8.9 略加修改，使用 1988 年通过选举而当选的现任总统在 1990 年的投票结果。候选人 A 于 1988 年当选而试图在 1990 年连任；$voteA90$ 表示候选人 A 在 1990 获得两党投票的份额。候选人 A 在 1988 年所获得的投票份额被用做候选人质量的代理变量。所有其他变量都是 1990 年选举中所出现的变量。利用 VOTE2.RAW 中的数据，估计了如下方程

$$\begin{aligned}\widehat{voteA90} = {} & \underset{(9.25)}{75.71} + \underset{(0.046)}{0.312}\,prtystrA + \underset{(1.01)}{4.93}\,democA - \underset{(0.684)}{0.929}\log(expendA)\\ & - \underset{(0.281)}{1.950}\log(expendB)\\ & n = 186, R^2 = 0.495, \bar{R}^2 = 0.483\end{aligned}$$

和

$$\begin{aligned}\widehat{voteA90} = {} & \underset{(10.01)}{70.81} + \underset{(0.052)}{0.282}\,prtystrA + \underset{(1.06)}{4.52}\,democA\\ & - 0.839\log(expendA) - 1.846\log(expendB) + 0.067\,voteA88\end{aligned}$$

$$(0.687) \qquad (0.292) \qquad (0.053)$$

$$n = 186, R^2 = 0.499, \bar{R}^2 = 0.485$$

（i）解释 *voteA*88 的系数并讨论其统计显著性。

（ii）添加 *voteA*88 对其他系数具有很大影响吗？

9.3 令 *math*10 表示密歇根州高中学生在一次标准化数学考试中的及格百分比（也可参见例 4.2）。我们感兴趣的是，平均对每个学生的支出对其成绩的影响。一个简单的模型是

$$math10 = \beta_0 + \beta_1 \log(expend) + \beta_2 \log(enroll) + \beta_3 poverty + u$$

其中 *poverty* 表示贫困生的比例。

307 （i）变量 *lnchprg* 表示学校有资格享受联邦政府午餐支助计划的学生比例。为什么它是 *poverty* 的一个合适的代理变量？

（ii）下表包含了有和没有 *lnchprg* 作为解释变量时的 OLS 估计值。解释为什么支出对 *math*10 的影响在第（2）列比在第（1）列要低。第（2）列中的这种影响在统计上仍大于 1 吗？

因变量：*math*10

自变量	(1)	(2)
log(*expend*)	11.13	7.75
	(3.30)	(3.04)
log(*enroll*)	0.022	1.26
	(0.615)	(0.58)
lnchprg	—	−0.324
		(0.036)
截距项	−69.24	−23.14
	(26.72)	(24.99)
观测次数	428	428
R-平方	0.029 7	0.189 3

（iii）在其他条件相同的情况下，越大的学校通过率越低吗？请解释。

（iv）解释第（2）列中 *inchprg* 的系数。

（v）你如何理解 R^2 从第（1）列到第（2）列的显著提高？

9.4 如下方程用儿童年龄、母亲受教育水平、父亲受教育水平和家庭子女数来解释儿童每周看电视的小时数：

$$tvhours^* = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 motheduc + \beta_4 fatheduc + \beta_5 sibs + u$$

我们担心在调查中有测量误差。令 *tvhours* 表示所报告的每周看电视的小时数。

(i) 在这个应用研究中，经典的含变量误差（CEV）假定有什么要求？

(ii) 你认为 CEV 假定可能成立吗？请解释。

9.5 在例 4.4 中，我们针对一个大学样本估计了一个联系校园犯罪与学生注册人数的模型。由于很多学校在 1992 年都没有报告其校园犯罪数据，308 所以我们所使用的样本并不是美国大学的一个随机样本。你认为大学没有报告校园犯罪可被视为外生的样本选择问题吗？请解释。

计算机习题

9.6 (i) 在习题 7.13 所估计的模型中应用方程（9.3）中的 RESET。此方程中有函数形式误设的证据吗？

(ii) 计算一个异方差稳健形式的 RESET。你在第（i）部分的结论改变了吗？

9.7 本题使用 WAGE2.RAW 中的数据。

(i) 在例 9.3 中，用变量 KWW（“工作领域内知识”考试分数）取代 IQ 作为能力的代理变量。在此情形下，估计的教育回报是多少？

(ii) 现在用 IQ 和 KWW 一起作为代理变量。所估计的教育回报会怎么样？

(iii) 在第（ii）部分中，IQ 和 KWW 是个别显著的吗？它们联合显著吗？

9.8 本题使用 JTRAIN.RAW 中的数据。

(i) 考虑简单回归模型

$$\log(scrap)=\beta_0+\beta_1 grant+u$$

式中，$scrap$ 为企业的废弃率；$grant$ 为表示是否得到在职培训津贴的一个虚拟变量。你能想到 u 中观测不到的因素可能会与 $grant$ 相关的原因吗？

(ii) 利用 1988 年的数据估计这个简单的回归模型。（你应该有 54 个观测。）得到在职培训津贴显著地降低企业的废弃率吗？

(iii) 现在增加一个解释变量 $\log(scrap_{87})$。这将如何改变 $grant$ 的估计影响？解释 $grant$ 的系数。相对单侧对立假设 H_1：$\beta_{grant}<0$，它在 5% 的显著性水平上统计显著吗？

(iv) 相对双侧对立假设检验 $\log(scrap_{87})$ 的参数为 1 的虚拟假设。报告检验的 p 值。

(v) 利用异方差—稳健性标准误重复第（iii）和第（iv）步，并简要讨论任何明显的差异。

9.9 本题使用 INFMRT.RAW 中 1990 年的数据。

(i) 重新估计方程（9.37），但现在对哥伦比亚特区这个观测引进一个虚拟变量（记为 DC）。解释 DC 的系数，并评论其大小和显著性。

(ii) 将第（i）步所得到的估计值和标准误与方程（9.38）中的估计值和标准误相比较。根据这种对单一观测引进一个虚拟变量的做法，你得到什么结论？

309 9.10 利用 RDCHEM.RAW 中的数据进一步考查异常数据对 OLS 估计值的影响。具体而言，就是在包括和不包括年销售额近 400 亿美元的企业的情况下，估计模型

$$rdintens = \beta_0 + \beta_1 sales + \beta_2 sales^2 + \beta_3 profmarg + u$$

并讨论结果在重要的方面是否不同。如果在估计前重新定义 *sales* 的度量单位为 10 亿美元，这个方程读起来就容易些（参见习题 6.3）。

9.11 去掉例 4.10 中教师津贴低于薪水 1%的学校，重新做这个例子。

(i) 这样做将减少多少个观测?

(ii) 去掉这些观测对所估计的取舍关系有重要影响吗?

9.12 本题使用 LOANAPP.RAW 中的数据。

(i) 有多少个观测的 $obrat > 40$，即其他债务负担超过其总收入的 40%?

(ii) 在习题 7.16 中，去掉 $obrat > 40$ 的观测，重新估计第（iii）部分中的模型。*white* 的系数估计值和 t 统计量将会怎样?

(iii) β_{white}看起来对所使用的样本过度敏感吗?

经济科学译丛·计量经济学导论·现代观点　经济科学译丛·计量经济学导论·现代观点　经济科学译丛·计量

第2篇

时间序列数据的回归分析

310 在对如何应用多元回归模型处理横截面数据问题有了清楚的了解之后，我们接下来学习时间序列数据计量经济学。我们将主要借助普通最小二乘法，而与之有关的具体操作和推断的工作我们在前面章节已经做过了。然而，正如我们在第1章所说，时间序列数据具有某些横截面数据没有的特点，在应用OLS时应该对这些特点给予特别的注意。

第10章包括基本的时间序列回归分析，注意到了时间序列数据所特有的问题。我们提出了用于时间序列数据的高斯-马尔科夫和经典线性模型的一系列假定。我们也讨论了函数形式、虚拟变量以及趋势和季节性等问题。

时间序列模型违背了高斯-马尔科夫假定，第11章描述了这些违背情况的性质，并提出了普通最小二乘法的大样本性质。由于不能继续保留随机抽样的假定，我们必须研究能够限制时间序列中**在时间上相关**(temporal correlation)的条件，以确保通常的渐近分析有效。第12章讨论了一个重要的新问题：时间序列回归中误差项的序列相关。我们讨论了这种相关的**后果**(consequences)、检验方法和补救的办法等议题。这一章也解释了在时间序列模型中异方差是如何产生的。

第 10 章　时间序列数据的基本回归分析

311　在这一章，我们开始接触使用时间序列数据的线性回归模型，并研究用于估计这种模型的 OLS 的性质。10.1 节讨论了时间序列数据和横截面数据的一些概念性区别。10.2 节提供了几个在经验社会科学中经常用到的时间序列回归的例子。接下来，我们把注意力转向 OLS 估计量的**有限样本性质**（finite sample properties），并介绍时间序列回归条件下的高斯-马尔科夫假定和经典线性模型假定。尽管这些假定与横截面情况下的假定有某些相同之处，但也存在着需要阐明的重要区别。

接着，我们回到在横截面回归中曾经遇到过的一些问题，诸如如何应用和解释对数函数形式和虚拟变量等问题。10.5 节讨论在多元回归中如何把趋势包括进来以及如何表明季节性的重要问题。

10.1 时间序列数据的性质

使时间序列数据区别于横截面数据的一个明显的特点是，时间序列数据集是按照时间排列顺序的。例如，在第 1 章我们简单地讨论了一系列时间序列数据，它们是反映波多黎各的就业、最低工资及其他经济变量的状况的数

据。我们清楚地看到，在这个数据组中，1971 年的数据紧跟在 1970 年的数据后面。在社会科学中，为了分析时间序列数据，我们必须确信过去可以影响未来，而不是相反。为了强调时间序列数据的恰当排序，表 10.1 列出 PHILLIPS.RAW 中美国通货膨胀和失业率的部分数据。

横截面数据和时间序列数据间的另一个区别更加微妙。我们在第 3 章和第 4 章研究了样本是随机地从适当的总体中抽出的情况下，OLS 估计量所具有的统计性质。对为什么横截面数据应该被视为随机结果不难理解：从总体中抽取的不同样本，通常会使得自变量和因变量（如教育、经验、工资等）有不同的取值。因此，通过不同的随机抽样计算出的 OLS 估计值通常不一样，这就是我们把 OLS 统计量当做随机变量的原因。

我们应该怎样认识时间序列数据的随机性呢？很明显，经济的时间序列满足随机变量的结果所要求的直观条件。例如，我们无法知道道琼斯工业指数在下一个交易日收盘时会是多少，我们也不知道加拿大下一年的年产出增长会是多少。既然这些变量的结果无法事先预料，它们当然应该被当做随机变量。

标注有时间的一个随机变量序列被正式地称做**随机过程**（stochastic process）或者**时间序列过程**（time series process）。当我们搜集到一组时间序列数据时，我们得到了该随机过程的一个可能的结果或实现（realization）。我们只能看到这一个实现，因为我们不能让时间倒转重新开始这个过程（这与横截面分析中只能搜集一个随机样本相似）。然而，如果过去的条件变成另外一种情况，我们通常会得到这个随机过程的另一种不同的实现，这正是把时间序列数据当成一系列随机变量的实现的原因。时间序列随机过程的所有可能的实现的组合相当于横截面分析中的总体。

312

表 10.1　　美国 1948—1996 年通货膨胀率和失业率部分数据

年份	通货膨胀率	失业率
1948	8.1	3.8
1949	1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
⋮	⋮	⋮
1994	2.6	6.1
1995	2.8	5.6
1996	3.0	5.4

10.2 时间序列回归模型的例子

在这一节，我们讨论时间序列模型的两个例子，它们在经验的时间序列分析中很有用，而且很容易用普通最小二乘法来估计。我们将在第 11 章研究其他模型。

静态模型

313 假使我们有两个变量（例如 y 和 z）的时间序列数据，并对 y_t 和 z_t 标注相同的时间。把 y 和 z 联系起来的一个静态模型为：

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad t = 1,2,\cdots,n \tag{10.1}$$

“静态模型”的名称来源于我们构造反映 y 和 z 之间同一时期的关系模型这一事实。当 z 在时间 t 的一个变化被认为对 y 有立即的影响时，通常设定一个静态模型，如，$\Delta y_t = \beta_1 \Delta z_t$，此时 $u_t = 0$。当我们对了解 y 和 z 之间的替换关系感兴趣时也使用静态回归模型。

一个静态模型的例子是**静态菲利普斯曲线**（static Phillips curve），表示为

$$inf_t = \beta_0 + \beta_1 unem_t + u_t \tag{10.2}$$

式中，inf_t 为年通货膨胀率；$unem_t$ 为失业率。这种形式的菲利普斯曲线实际上假定了一个不变的自然失业率和固定的通货膨胀预期。它可以用来研究同一时期内失业率和通货膨胀之间的相互替换关系。[参见 Mankiw（1994，第 11 章 2 节）]

当然，在一个静态回归模型中也可以有几个解释变量。令 $mrdrte_t$ 代表某个特定的城市在第 t 年中平均每千人中发生的谋杀案件数，$convrte_t$ 代表谋杀案的定罪率，$unem_t$ 代表本地失业率，$yngmle_t$ 代表当地人口中年龄在 18 岁～25 岁之间的男性人口的比例。这样，一个用来解释谋杀案发生率的静态多元回归模型为：

$$mrdrte_t = \beta_0 + \beta_1 convrte_t + \beta_2 unem_t + \beta_3 yngmle_t + u_t \tag{10.3}$$

借助这样的模型，我们希望能够估计出在其他变量保持不变时定罪率的增长对犯罪活动的影响，以及类似的问题。

有限分布滞后模型

在**有限分布滞后模型**［finite distributed lag（FDL）model］中，我们容

许一个或多个变量对 y 的影响有一定时滞。例如，考察如下模型：

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t \qquad (10.4)$$

式中，gfr_t 为生育率（每 1 000 个育龄妇女生育孩子个数）；pe_t 为个人税收豁免的实际美元金额。我们想从总体上来看一看，生育孩子的决策是否与因生孩子所放弃的税收豁免的价值有关系。方程（10.4）意味着，由于生理的和行为上的原因，个人税收豁免的变化并不马上就影响生育孩子的决定。

方程（10.4）是下面这个模型的一个例子：

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \qquad (10.5)$$

314 它是一个二阶 FDL。为了理解式（10.5）中的系数，设想 z 在 t 之前的所有时间里都是一个等于 c 的常数，在时间 t，z 增大一个单位变为 $c+1$，然后在时间 $t+1$ 再回到原来的水平。（即 z 的变大是暂时的。）准确地表示为

$$\cdots, z_{t-2}=c, z_{t-1}=c, z_t=c+1, z_{t+1}=c, z_{t-2}=c, \cdots$$

为集中研究其他条件不变时 z 对 y 的作用，设每个时期的误差项均为零。那么，

$$\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0 (c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0 c + \delta_1 (c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 (c+1) \\
y_{t+3} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

从前面两个式子得到 $y_t - y_{t-1} = \delta_0$，它表明 δ_0 是 z 在 t 时期增大一个单位所引起的 y 的即期变化。δ_0 通常被称做**即期倾向**（impact propensity）或**即期乘数**（impact multiplier）。

类似地，$\delta_1 = y_{t+1} - y_{t-1}$ 是 z 的这个暂时变化后的下一个时期 y 的变化，$\delta_2 = y_{t+2} - y_{t-1}$ 是这个变化后第二个时期的 y 的变化。在时间 $t+3$，y 回到了原来的水平：$y_{t+3} = y_{t-1}$，这是因为我们假定式（10.5）中 z 只有两期滞后。如果把 δ_j 当成 j 的函数，并作成一张表，就得到了**滞后分布**（lag distribution），它概括了 z 的一个暂时的变化对 y 的动态影响。图 10.1 给出了二阶 FDL 模型的一个可能的滞后分布。（当然，我们永远不知道参数 δ_j 的真实值，只能估计出 δ_j 的值，然后再画出估计的滞后分布。）

图 10.1 中的滞后分布暗示了最大的影响发生在第一期滞后。这个滞后分布有一种很有用的解释。如果我们把 y 的初始值标准化为 $y_{t-1}=0$，这个滞后分布就会描绘出 z 的一个单位的暂时增长所引起的 y 的随后所有取值）。

我们也对 z 的持久增长所引起的 y 的变化感兴趣。在时间 t 之前，z 等于常数 c。从时间 t 起，z 永远地增大为 $c+1$。即当 $s<t$ 时，$z_s=c$；当 $s \geqslant t$ 时，$z_s=c+1$。我们再把误差项都设为零，得到

$$y_{t-1} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c$$

$$y_t = \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c$$
$$y_{t+1} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c$$
$$y_{t+2} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1)$$
…………

315 随着 z 从时间 t 起永久地变大，在其后的一个时期 y 增大了 $\delta_0+\delta_1$，其后的两个时期 y 增大了 $\delta_0+\delta_1+\delta_2$。两个时期后，$y$ 没有进一步的变化。这表明，当期的和滞后的 z 的系数之和 $\delta_0+\delta_1+\delta_2$ 等于 z 的持久性增长引起的 y 的长期变化，它被称为**长期倾向**（LRP）或**长期乘数**。LRP 是在分布滞后模型中人们经常关注的问题。

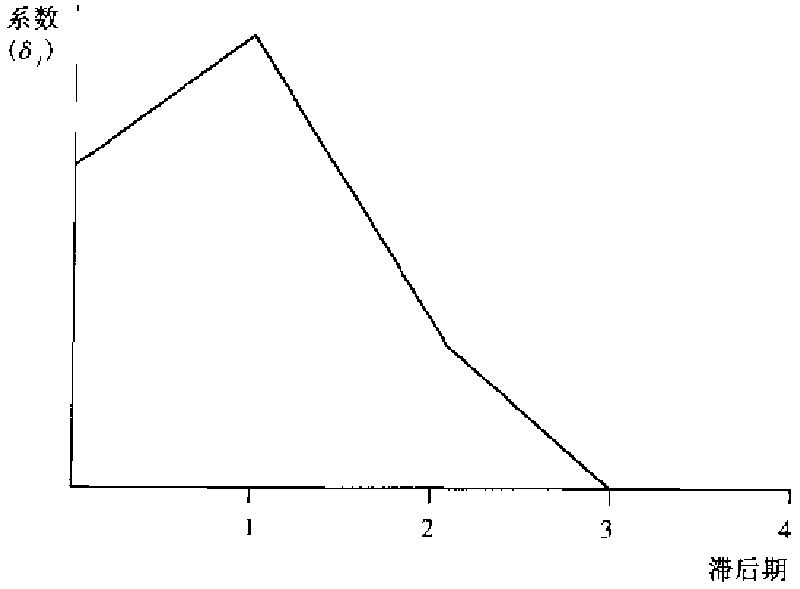


图 10.1　一个有两期非零滞后的滞后分布：最大的影响发生在第一期滞后

举例来说，在方程（10.4）中，δ_0 表示出了 pe 增长 1 美元带来的生育率的当期变化，正如我们以前所言，有理由相信 δ_0 如果不是零就是一个比较小的数。但是，δ_1 和 δ_2 二者之一或二者都会是正的。如果 pe 永久性地增加 1 美元，两年之后，gfr 有 $\delta_0+\delta_1+\delta_2$ 这么大幅度的变化。这个模型假定两年后没有进一步的变化，这一点是否符合实际是一个经验问题，这里不作讨论。

一个 q 阶有限分布滞后模型表示为

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + u_t \tag{10.6}$$

静态模型作为一种特例也被包括在内了，只要把 δ_1，δ_2，…，δ_q 设为零就行了。有时，估计一个滞后分布模型的主要目的是检验 z 是否对 y 有滞后的影响。即期倾向总是当期 z 的系数 δ_0；我们偶尔从方程（10.6）中把 z_t 省略掉，这样，即期倾向为零，滞后分布又被描绘成了 j 的函数 δ_j。长期影响是变量 z_{t-j} 的系数的总和：

$$\mathrm{LRP} = \delta_0 + \delta_1 + \cdots + \delta_q \tag{10.7}$$

316 因为 z 在不同时期的滞后之间经常有较大程度的相关，也就是，方程 (10.6) 中存在多重共线性，我们很难准确地估计出单独一个 δ_j。有意思的是，即使在无法准确估计 δ_j 时，我们也经常能很好地估计出 LRP。我们将在后面举例说明。

问题 10.1 对一组年度数据，假定相应的方程为

$$int_t = 1.6 + 0.48inf_t - 0.15inf_{t-1} + 0.32inf_{t-2} + u_t$$

式中，int 为利率；inf 为通货膨胀率。那么，即期倾向和长期倾向分别是多少？

我们可能遇到不止一个带有滞后的变量，也可能需要在 FDL 模型中添加更多的当期变量。例如，育龄妇女的平均教育水平可以作为一个变量加入到式 (10.4) 中去，以说明改变妇女教育水平的作用。

标注时间的惯例

当模型中含有滞后的解释变量时（下一章我们会接触到带有滞后的 y），在如何标注初始观测值方面容易产生混乱。比如，假定方程 (10.5) 是对的，如果假设观测从 $t=1$ 开始，那么在第一个时期解释变量是 z_1，z_0 和 z_{-1}。我们的惯例将是：既然它们是样本中的初始值，我们就从 $t=1$ 开始标注时间。实际上，这不是什么重要问题，因为回归分析软件包会自动记录可以用来估计滞后模型的所有观测。但是在本章和下一章我们有必要对回归方程中第一个时期的表达习惯作出规定。

10.3 经典假设下 OLS 的有限样本性质

这一节，我们将完整地列出标准假定下 OLS 的有限样本（小样本）性质。我们要特别注意，为了把时间序列回归的情形包括进去，必须对那些假定作出怎样的修改。

OLS 的无偏性

第一个假定指时间序列过程符合对参数为线性的模型。

假定 TS.1（对参数是线性的）

随机过程 $\{(x_{t1}, x_{t2}, \cdots, x_{tk}, y_t): t = 1, 2, \cdots, n\}$ 遵循线性模型

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \qquad (10.8)$$

式中，$\{u_t: t = 1, 2, \cdots, n\}$ 为误差或干扰项序列。这里，n 是观测的次数（时期）。

317 符号 x_{tj} 中，t 表示时期，j 和平常一样表示 x_{tj} 是 k 个解释变量中的一个。横截面回归中使用的术语在这里同样有效：y_t 是因变量、被解释变量或回归子；x_{tj} 是自变量、解释变量或回归元。

我们应该把假定 TS.1 看做与假定 MLR.1（横截面回归的第一个假定）本质上是相同的，现在只不过把线性模型具体应用于时间序列数据。第二节中的例子可以写成式（10.8）的形式，只要适当地定义 x_{tj} 就行了，例如，设 $x_{t1} = z_t$，$x_{t2} = z_{t-1}$，$x_{t3} = z_{t-2}$，就可以得出方程（10.5）。

为了便于阐述和讨论余下的几个假定，我们令 $\mathbf{x}_t = (x_{t1}, x_{t2}, \cdots, x_{tk})$ 代表方程中在时间 t 上的自变量集，令 $\mathbf{X}$ 代表所有时期内全部自变量集。不妨把 $\mathbf{X}$ 看做由 n 行和 k 列组成的阵列，它反映了在经济软件包中时间序列数据的存储方式：$\mathbf{X}$ 的第 t 行为 $\mathbf{x}_t$，由时期 t 的所有独立变量组成。因此，$\mathbf{X}$ 的第一行与 $t=1$ 对应，第二行与 $t=2$ 对应，第 n 行与 $t=n$ 对应。表 10.2 给出了一个例子，它使用了方程（10.3）中的解释变量，并取 $n=8$。

表 10.2　一个例子：方程（10.3）的诸解释变量 X

t	*convrte*	*unem*	*yngmle*
1	0.46	0.074	0.12
2	0.42	0.071	0.12
3	0.42	0.063	0.11
4	0.47	0.062	0.09
5	0.48	0.060	0.10
6	0.50	0.059	0.11
7	0.55	0.058	0.12
8	0.56	0.059	0.13

时间序列的下一个假定与假定 MLR.3 类似，它去掉了假定 MLR.2 中随机抽样的假定。

假定 TS.2（条件均值为零）

对于每一个 t，当所有时期的解释变量都给定时，误差项 u_t 的期望值为零。用数学表示为

318

$$E(u_t|\mathbf{X})=0,\quad t=1,2,\cdots,n \tag{10.9}$$

这是个非常重要的假定，我们需要对它的含义有个直观的理解。在横截面的情况下，从非相关性的角度来考虑这个假定非常容易：假定 TS.2 意味着时间为 t 的误差项 u_t 与任何时期的解释变量都线性无关。我们是用条件均值来阐述这个假定的，这一事实意味着我们必须正确地设定 y_t 和解释变量之间的函数关系。如果 u_t 独立于 $\mathbf{X}$ 且 $E(u_t)=0$，那么假定 TS.2 自动成立。

了解了第 3 章的横截面分析，我们就不会感到奇怪，在时间序列的情况下需要 u_t 与时间下标为 t 的解释变量不相关：用条件均值来表示，即

$$E(u_t|x_{t1},\cdots,x_{tk})=E(u_t|\mathbf{x}_t)=0 \tag{10.10}$$

当式（10.10）成立时，我们称 x_{tj} 是**同期外生**（contemporaneously exogenous）的。方程（10.10）暗示了 u_t 和解释变量之间是同期不相关的：$\text{Corr}(x_{tj},u_t)=0$，对所有的 j 成立。

假定 TS.2 要求的不仅仅是同期外生：u_t 必须与 x_{sj} 不相关，即使 $s\neq t$。这是在很强的意义上说解释变量是外生的。因此，当 TS.2 成立时，我们称解释变量是**严格外生**（strictly exogenous）的。在第 11 章，我们将阐述式（10.10）足以保证 OLS 估计量的一致性。但是，为了表明 OLS 是无偏的，我们需要严格外生性假定。

在横截面的情况下，我们并没有明确地指出某个人 i 的误差项与样本中其他人的解释变量的关系如何。没有必要那样做的原因是，由于抽样是随机的（假定 MLR.2），u_i 自动地与第 i 次以外的其他观测中的解释变量保持独立的关系。在时间序列的情况下，把抽样当成随机的是很不恰当的，因此，我们必须明确假定 u_t 的期望值在任何时期都与解释变量不相关。很重要的一点是，假定 TS.2 并没有对自变量之间或不同时间的 u_t 的相关性作出任何限制，它只提到 u_t 的平均值在任何时期都与解释变量无关。

能引起时间 t 的未观测到的变量与解释变量在任何时期都有相关关系的行为，都会使假定 TS.2 无效。导致无效的两个主要的可能性是遗漏变量和对某些回归元的测量误差。但严格外生性假定也可能因为其他更不明显的原因而变为无效。在简单的静态模型中

$$y_t=\beta_0+\beta_1z_t+u_t$$

假定 TS.2 不但要求 u_t 和 z_t 不相关，而且要求 u_t 与过去的和将来的 z 都不相关。这有两重含义。第一，z 对 y 没有滞后的影响，如果 z 的确对 y 有滞后的影响，那么我们需要估计一个分布滞后模型。更微妙的一点是，严格外生性排除了现在的误差项的变化引起将来 z 变化的可能性，这事实上否定

319 了 y 对 z 的将来值的反馈作用。举个例子，我们考虑一个简单的用人均警官数来解释一个城市的犯罪率的静态模型：

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t$$

假定 u_t 与 $polpc_t$ 甚至过去 $polpc_t$ 的值不相关也许有道理；为了方便讨论，我们假定事实就是如此。但是，设想这个城市根据过去犯罪率的多少来调整警力规模，这就意味着 $polpc_{t+1}$ 可能会与 u_t 相关（因为较高的 u_t 导致较高的 $mrdrte_t$）。如果事实确实是这样，假定 TS.2 一般不成立。

在分布滞后模型中，也存在这样的顾虑。通常，我们不担心 u_t 会与过去的 z 相关，因为在模型中控制了过去的 z。u 对将来的 z 的反馈却是一个令人头痛的问题。

严格外生的解释变量无法对过去发生在 y 上的变化作出反应。农业生产函数中的降雨量之类的因素满足这样的要求：将来各年的降雨量不受现在或过去产量的影响。但是，类似于劳动投入量这种变量可能不是严格外生的，因为它是由农民决定的，而农民可能根据上一年的产量调整劳动投入。政策变量，如货币供给的增长、福利开支、高速公路速度限制等经常受结果变量过去的情况的影响。在这些社会科学中，很多解释变量会违背严格外生性假定。

即使 TS.2 不太现实，我们还是要以它为出发点，目的是得出 OLS 估计量具有无偏性的结论。大多数处理静态和有限分布滞后模型的方法为了确保 TS.2 成立，作出解释变量不是随机的或在重复抽样中保持不变这一更强的假定。对于时间序列观测而言，非随机性假定显然是错误的；假定 TS.2 具有更符合 x_{tj} 有随机性质的优点。同时，为了使 OLS 无偏，它把对 u_t 和解释变量的关系所作的必要假定分离了出来。

OLS 无偏性要求的最后一个假定是标准的无完全的多重共线性假定。

假定 TS.3（无完全多重共线性）

在样本中（并因此也在潜在的时间序列过程中），没有任何自变量是恒定不变的，或者是其他自变量的一个完全的线性组合。

在第 3 章中我们详细讨论过横截面数据条件下的这个假定。这个问题在时间序列的条件下本质上是一样的。记住，假定 TS.3 容许解释变量之间的相关，但它不容许样本中的完全相关。

定理 10.1（OLS 的无偏性）

当假定 TS.1，TS.2 和 TS.3 成立时，对于给定的 $\mathbf{X}$，OLS 的估计量是无偏的，从而 $\mathrm{E}(\hat{\beta}_j) = \beta_j (j = 0, 1, \cdots, k)$ 也无条件地成立。

这个定理的证明与第 3 章的定理 3.1 的证明实质上是相同的，因此这里将它省略。把定理 10.1 与定理 3.1 进行比较发现，我们放弃了随机抽样的假定，取而代之的是，假定给定所有时期的解释变量时，对每个时间 t，u_t 的均值均为零。如果这个假定不成立，则不能证明 OLS 是无偏的。

问题 10.2 320 在 FDL 模型 $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + u_t$ 中，我们需要对序列 z_0，z_1，…，z_n 作怎样的假定才能使假定 TS.3 成立呢？

3.3 节对遗漏变量产生偏误的分析，在时间序列的情况下同样适用。特别地，表 3.2 及围绕它展开的讨论可以像以前那样，用来决定处理由于遗漏变量而产生的偏误的办法。

OLS 估计量的方差和高斯-马尔科夫定理

我们需要增加两个假定才能完成时间序列回归的高斯-马尔科夫定理。其中的第一个假定我们在横截面分析中曾遇到过。

假定 TS.4（同方差性）

给定 $\boldsymbol{X}$，u_t 的方差在所有的时间 t 上都相等：

$$\mathrm{Var}(u_t|\boldsymbol{X}) = \mathrm{Var}(u_t) = \sigma^2,\quad t = 1,2,\cdots,n$$

这个假定的意思是$\mathrm{Var}(u_t|\boldsymbol{X})$不能依赖于 $\boldsymbol{X}$——只要 u_t 和 $\boldsymbol{X}$ 相互独立就足够了——而且，$\mathrm{Var}(u_t)$在所有的时期里都是恒定的。当 TS.4 不成立时，跟横截面的情形一样，我们称误差是异方差的。这方面的例子很多，比如，一个决定 3 个月国债利率（$i3_t$）的方程

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t \tag{10.11}$$

式中，inf_t 为通货膨胀率；def_t 为联邦赤字占国内生产总值的百分比。TS.4 要求观测不到的、影响利率的因素在不同的时期有不变的方差。因为政策的变更会影响利率的方差，所以这个假定很可能是不对的；另外一种可能的情况是，利率的方差依赖于通货膨胀水平或赤字的相对规模，这也会违背同方差性假定。

当$\mathrm{Var}(u_t|\boldsymbol{X})$确实依赖于 $\boldsymbol{X}$ 时，它常常是依赖于在时间 t 上的解释变量 $\boldsymbol{x}_t$。在第 12 章中，我们将看到第 8 章中异方差的检验方法可以用在时间序列回归中，至少在一定的假设条件下是这样的。

时间序列分析的最后一个高斯-马尔科夫假定是以前未曾接触过的。

假定 TS.5（无序列相关）

给定 $\boldsymbol{X}$，任意两个不同时期的误差不相关：

$$\mathrm{Corr}(u_t,\ u_s|\boldsymbol{X}) = 0,\quad t \neq s$$

321 理解这个假定的最简单的方法是忽略它是以 $\boldsymbol{X}$ 为条件的。这样，假定 TS.5 变成

$$\mathrm{Corr}(u_t, u_s) = 0,\quad t \neq s \tag{10.12}$$

（这也是当 $\mathbf{X}$ 被当成非随机时表述无序列相关假定的方式。）在考察假定 TS.5 是否成立时，我们把注意力集中在方程（10.12）上，因为它容易理解一些。

当方程（10.12）不成立时，我们称方程（10.8）中的误差是**序列相关**（serial correlation）的或**自相关**（autocorrelation）的，这是因为误差在不同时期相关。考虑相邻时期误差的情形：当 $u_{t-1}>0$ 时，一般来说，下一个时期的误差 u_t 也是正的，那么就有 $\mathrm{Corr}(u_t,u_{t-1})>0$，这样，误差项是序列相关的——在方程（10.11）中它的含义是，如果当前的利率意外地高，那么下一时期的利率就可能会高于平均水平（对既定的通货膨胀和赤字水平而言）。我们将在第 12 章看到，这是对许多时间序列应用中的误差项的合理描述。现在，我们就假定 TS.5 成立。

很重要的是，假定 TS.5 没有提到自变量之间在时间上相关的问题。例如，方程（10.11）中，inf_t 在不同的时期几乎必然是相关的，但这与 TS.5 成立与否无关。

一个很自然的问题是：为什么在第 3 章和第 4 章中我们不假定不同的横截面观测值的误差是不相关的呢？答案在于随机抽样的假定：当抽样是随机时，对于任意两次观测 i 和 h，u_i 和 u_h 是相互独立的。在给定样本中的所有解释变量的条件下，可以证明它的正确性。因此，就我们当前的目的而言，序列相关是时间序列回归中特有的问题。

假定 TS.1～TS.5 是应用于时间序列的高斯-马尔科夫假定，但它们也有其他用途。有时，TS.1～TS.5 能在横截面应用中得到满足，即使是在随机抽样不是合理的假定时——比如在当横截面单元（units）相对于总体来说够大时就是如此；再比如，同一个州之内的不同城市之间可能是相关的，但只要各个城市的误差之间不相关，假定 TS.5 就成立。然而，我们的主要兴趣还是要把这些假定应用在时间序列回归模型中。

定理 10.2（OLS 的样本方差）

在时间序列的高斯-马尔科夫假定 TS.1～TS.5 成立时，$\hat{\beta}_j$ 对 $\mathbf{X}$ 的条件方差为

$$\mathrm{Var}(\hat{\beta}_j|\mathbf{X})=\sigma^2/[\mathrm{SST}_j(1-R_j^2)],\ j=1,\cdots,k \tag{10.13}$$

式中，SST_j 为 x_{tj} 的总的平方和；R_j^2 为由 x_j 对其他自变量回归得到的 R^2。

322 式（10.13）中的方差与第 3 章得出的横截面高斯-马尔科夫假定下的完全一样。既然它的证明与定理 3.2 的证明非常相似，这里将其省略。第 3 章中关于包括解释变量间的多重共线性在内的各种造成大方差的因素的讨论，都可以直接应用到时间序列分析中。

在假定 TS.1～TS.5 下，通常的误差方差的估计量是无偏的，高斯-马尔科夫定理有效。

定理 10.3（$\hat{\sigma}^2$ 的无偏估计）

在假定 TS.1－TS.5 下，估计量 $\hat{\sigma}^2 = \text{SSR}/df$ 是 σ^2 的一个无偏估计量，式中，$df = n - k - 1$。

定理 10.4（高斯-马尔科夫定理）

在假定 TS.1－TS.5 下，给定 **X** 的值，OLS 估计量是最优线性无偏估计。

问题 10.3 在 FDL 模型 $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + u_t$ 中，试说明解释变量中任何多重共线性的性质。

归结起来，OLS 在假定 TS.1－TS.5 下和假定 MLR.1－MLR.5 下具有同样理想的有限样本性质。

经典线性模型假定下的推断

为了能够使用通常的 OLS 标准误、t 统计量和 F 统计量，我们需要增加最后一个假定，它类似于横截面分析中的正态性假定。

假定 TS.6（正态性）

误差 u_t 独立于 **X**，且与 Normal（0，σ^2）是独立同分布的。

假定 TS.6 蕴涵了 TS.3，TS.4 和 TS.5，但它更强，因为它还假定了独立性和正态性。

定理 10.5（正态抽样分布）

在时间序列的 CLM 假定 TS.1－TS.6 下，给定 **X**，OLS 估计量遵循正态分布。而且，在虚拟假设下，每个 t 统计量呈 t 分布，F 统计量呈 F 分布，通常构造置信区间的方法也是有效的。

323 定理 10.5 的内涵非常重要。它告诉我们，当假定 TS.1－TS.6 成立时，我们已经学过的关于横截面回归的估计和推断的全部内容都可以直接用到时间序列回归中。这样，t 统计量可以用来检验个别解释变量的统计显著性，而 F 统计量可用来检验联合显著性。

正如在横截面情况下一样，推断方法的优劣和隐含的假定的真伪是一致的。经典线性模型用在时间序列数据上比用在横截面数据上要受到更多的限制——特别是，严格外生性和无序列相关的假定不太现实。尽管如此，CLM 框架对很多应用而言却是一个很好的起点。

例 10.1　静态菲利普斯曲线

为确定一般情况下失业和通货膨胀之间是否存在互相替换的关系，我们可以在方程（10.2）中检验 H_0：$\beta_1=0$，对立于 H_0：$\beta_1<0$。如果经典线性模型假定成立，我们可以使用通常 OLS 计算的 t 统计量。利用 PHILLIPS.RAW 中 1948—1996 年美国每年度数据资料，我们得到

$$\begin{aligned} \widehat{inf}_t &= 1.42 + 0.468\,unem_t \\ &\quad(1.72)\,(0.289) \qquad\qquad (10.14) \\ &\quad n=49,\ R^2=0.053,\ \bar{R}^2=0.033 \end{aligned}$$

这个方程并没有显示出 $unem$ 和 inf 之间有替换关系：$\hat{\beta}_1>0$。相反，如果通货膨胀和失业之间的确有什么关系的话，那也是正向关系。

以上的分析存在着一些问题，但我们现在还无法详细论述，在第 12 章我们将发现 CLM 假定并不成立。另外，静态菲利普斯曲线可能不是判断通货膨胀和失业之间是否有短期替代关系的最佳模型。宏观经济学家一般更倾向于使用附加预期的菲利普斯曲线，我们将在第 11 章举一个简单的例子。

作为第二个例子，我们利用美国经济的年度数据估计方程（10.11）。

例 10.2　通货膨胀和赤字对利率的影响

INTDEF.RAW 中的数据来自于 1997 年总统经济报告，时间从 1948 年到 1996 年。变量 $i3$ 是 3 个月国债利率，inf 是根据消费价格指数（CPI）得出的年通货膨胀率，def 是联邦赤字占 GDP 的百分比。所估计的方程为：

324

$$\begin{aligned} \widehat{i3}_t &= 1.25 + 0.613\,inf_t + 0.700\,def_t \\ &\quad(0.44)\,(0.076)\qquad(0.118) \qquad\qquad (10.15) \\ &\quad n=49,\ R^2=0.697,\ \bar{R}^2=0.683 \end{aligned}$$

估计结果表明，通货膨胀和赤字相对规模的增长一起作用于短期利率，使利率上升，它们对利率的影响与基本经济理论所预言的一致。如果通货膨胀率单独增加 1% 时，$i3$ 增长 0.613%。inf 和 def 都是统计上显著的，当然，这需要以 CLM 假定成立为前提条件。

10.4　函数形式、虚拟变量和指数

在前面章节中学到的所有函数形式都可以用在时间序列回归中。其中，最重要的是自然对数函数：在应用研究中经常出现具有恒定的百分比效应的

时间序列回归。

例 10.3　波多黎各的就业和最低工资

卡斯蒂罗-弗里德曼和弗里德曼（Castillo-Freedman and Freedman，1992）利用波多黎各的就业率、最低工资和其他变量的年度数据来研究美国的最低工资对波多黎各的就业的影响。一个简化了的模型为

$$\log(prepop_t)=\beta_0+\beta_1\log(mincov_t)+\beta_2\log(usgnp_t)+u_t \tag{10.16}$$

式中，$prepop_t$ 为波多黎各第 t 年的就业率（就业人口占总人口的比例）；$usgnp_t$ 为美国的实际国民生产总值（单位：十亿美元）；$mincov$ 度量着最低工资相对于平均工资的重要性，具体来说，$mincov=(avgmin/avgwage)\cdot avgcov$，其中，$avgmin$ 是平均最低工资，$avgwage$ 是总体平均工资，$avgcov$ 是平均工资覆盖率（实际上受益于最低工资法的工人的比例）。利用 1950—1987 年的数据，得出

$$\begin{aligned}\log(prepop_t)=&-\underset{(0.77)}{1.05}-\underset{(0.065)}{0.154}\log(mincov_t)-\underset{(0.089)}{0.012}\log(usgnp_t)\\&n=38,\ R^2=0.661,\ \bar{R}^2=0.641\end{aligned}\tag{10.17}$$

估计的 $prepop$ 对 $mincov$ 的弹性是 -0.154，而且，$t=-2.37$，统计上是显著的。因此，更高的最低工资降低了就业率，这与古典经济学的预言一样。GNP 变量统计上不显著，但是，下一节我们把时间趋势考虑进来时将得出不同的结论。

325　我们也可以把对数函数形式用于分布滞后模型中。例如，假定用季度数据表示的货币需求（M_t）与国内生产总值（GDP_t）之间的关系为

$$\begin{aligned}\log(M_t)=&\alpha_0+\delta_0\log(GDP_t)+\delta_1\log(GDP_{t-1})+\delta_2\log(GDP_{t-2})\\&+\delta_3\log(GDP_{t-3})+\delta_4\log(GDP_{t-4})+u_t\end{aligned}$$

方程中的即期倾向 δ_0 也被称为**短期弹性**(short-run elasticity)：它度量了 GDP 增长 1%时货币供给的当期变化的百分比。长期倾向 $\delta_0+\delta_1+\cdots+\delta_4$ 有时也被称为**长期弹性**(long-run elasticity)：它度量了 GDP 持久地增长 1%的四个月后的货币供给变化的百分比。

二元的或虚拟自变量在时间序列的应用中也非常有用。既然观测的单位是时间，虚拟变量代表在每个时期中某特定事件是否发生。例如，我们可以定义一个年度的虚拟变量 $democ_t$，用它表示在每一年中美国总统属于民主党还是共和党，如果是民主党，变量取值 1，否则取值 0。或者，在考察罚款对得州犯罪率的影响时，我们定义一个年度的虚拟变量，如果得州在那一年有罚款制度，变量值为 1，否则为 0。

虚拟变量经常被用来在数据集所涉及的多个时期中，把某些与其他时期有系统性区别的时期分离出来。

例 10.4　个人税收豁免对生育率的影响

总生育率（gfr）是每 1 000 个育龄妇女生育孩子的个数。对 1913—1984 年这段时间，方程

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 ww2_t + \beta_3 pill_t + u_t$$

用个人税收豁免的实际美元金额（pe）和两个虚拟变量解释了 gfr。变量 $ww2$ 在 1941 年和 1945 年间为 1，这个时期美国被卷入第二次世界大战。变量 $pill$ 自 1963 年后一直为 1，那时避孕药出现，开始被用来控制生育。

利用 FERTIL3.RAW 中的数据——它们来自惠廷顿、阿尔姆和彼德斯（Whittington，Alm and Peters，1990）的文章——得到

$$\begin{aligned} \hat{gfr}_t &= 98.68 + 0.083 pe_t - 24.24 ww2_t - 31.59 pill_t \\ &\quad (3.21)\ (0.030)\quad (7.64)\qquad (4.08) \qquad\qquad (10.18) \\ &\quad n = 72,\ R^2 = 0.473,\ R^2 = 0.450 \end{aligned}$$

每一个变量的双侧检验都是在 1% 水平上显著的。我们发现第二次世界大战期间的生育率较低：对于给定的 pe，gfr 减少了 24 个，降幅很大。（1913—1984 年，gfr 在 65～127 之间变动。）类似地，自从控制生育的药品产生以来，生育率也大幅下降。

326 有经济意义的变量是 pe。这段时期 pe 的平均值是 100.40 美元，变化范围是从 0～243.83 美元。pe 的系数表明 pe 每增加 12 美元 gfr 大约增加 1 个，这种影响不算太小。

在 10.2 节我们提到生育率可能受到 pe 变化的滞后作用。估计一个有两期滞后的分布滞后模型，得到

$$\begin{aligned} \hat{gfr}_t &= 95.87 + 0.073 pe_t - 0.005\,8 pe_{t-1} + 0.034 pe_{t-2} - 22.13 ww2_t \\ &\quad (3.28)\ (0.126)\quad (0.155\,7)\qquad (0.126)\qquad (10.73) \\ &\quad - 31.30 pill_t \\ &\quad (3.98) \qquad\qquad\qquad\qquad\qquad\qquad (10.19) \\ &\quad n = 70,\ R^2 = 0.499,\ R^2 = 0.459 \end{aligned}$$

在这个回归中，我们只有 70 次观测，这是因为 pe 滞后两次减少了 2 次观测。pe 的各项系数估计得很不准确，每一个变量单独看都不显著。事实上，pe_t，pe_{t-1} 和 pe_{t-2} 之间较大程度地相关，多重共线性的存在使得估计每个 pe 的系数非常困难。然而，pe_t，pe_{t-1} 和 pe_{t-2} 是联合显著的，F 统计量的 p 值为 0.012。因此，pe 的确对 gfr 有影响［正如我们在式（10.18）所见］，但我们还不能足够好地估计出影响是当期的还是有一期或两期的滞后（或它们中的几个）。实际上，pe_{t-1} 和 pe_{t-2} 是联合不显著的（p 值 $=0.95$），正是由于这个原因，我们应该改用静态模型。不过，我们不妨把它当做一个例子，来计算一下本模型中长期倾向的置信区间。

由式（10.19）估计得到的 LRP 是 $0.073 - 0.005\,8 + 0.034 \approx 0.101$。但我们无法从式（10.19）中得出这个估计值的标准误。为得到 LRP 估计值的

标准误，我们使用 4.4 节建议的技巧，令 $\theta_0 = \delta_0 + \delta_1 + \delta_2$ 代表 LRP，用 θ_0，δ_1 和 δ_2 表示 δ_0，$\delta_0 = \theta_0 - \delta_1 - \delta_2$。下一步，把模型 $gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + \cdots$ 中的 δ_0 替换掉，得到

$$\begin{aligned} gfr_t &= \alpha_0 + (\theta_0 - \delta_1 - \delta_2) pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + \cdots \\ &= \alpha_0 + \theta_0 pe_t + \delta_1 (pe_{t-1} - pe_t) + \delta_2 (pe_{t-2} - pe_t) + \cdots \end{aligned}$$

在这个方程中，我们可以通过用 pe_t，$(pe_{t-1} - pe_t)$，$(pe_{t-2} - pe_t)$，$ww2_t$ 和 $pill_t$ 做 gfr_t 的回归得出 $\hat{\theta}_0$ 和它的标准误。pe_t 的系数及其标准误就是我们需要的。进行回归得到 $\hat{\theta}_0 = 0.101$（与上面的结果一样）和 se（$\hat{\theta}_0$）$= 0.030$［从式（10.19）无法计算出来］。$\hat{\theta}_0$ 的 t 统计量为 3.37，$\hat{\theta}_0$ 在较小的显著性水平上是异于 0 的。即使每个 δ_j 单独地都不显著，但 LRP 非常显著，它的 95% 置信区间是从大约 0.041～0.160。

惠廷顿、阿尔姆和彼德斯（Whittington, Alm and Peters, 1990）考虑了更多期滞后的可能，但他们对系数作出了限制以消除会妨碍估计单个 δ_j 的多重共线性问题（具体方法参见例 10.6）。LRP 是我们的主要兴趣所在，为

327 了估计它，作这种限制是必要的。惠廷顿、阿尔姆和彼德斯还增加了女性平均工资和失业率等变量。

在所谓的**事件研究**（event study）中，虚拟变量是个关键的部分。事件研究的目标是确定某个特定的事件是否会影响到某项结果。研究产业组织的经济学家探究了特定事件对公司股票价格的影响。例如，罗斯（Rose, 1985）研究了新的运输管理条例对运输公司股票价格的影响。

用于这类事件研究的一个简单的方程是

$$R_t^f = \beta_0 + \beta_1 R_t^m + \beta_2 d_t + u_t$$

式中，R_t^f 为在时期 t（通常为 1 周或 1 个月）内来自于公司 f 的股票收益；R_t^m 为市场收益率（通常由整个股票市场的指数计算而得）；d_t 为用来表示某一事件发生时间的虚拟变量。比如，对于一个航空公司，d_t 可以表示该公司在第 t 周遇到了一个公开了的事件或将要发生的事件。把 R_t^m 包括在方程中，是因为存在着整个市场的运行与航空事件相对应的可能性。有时，在回归分析中也使用多个虚拟变量。比如，如果事件是指颁布了一项关系到某些公司的新管理条例，我们就要用一个虚拟变量代表条例公布前的几个星期，用另一个虚拟变量代表条例公布后的几个星期，其中，通过第一个虚拟变量可以推测出是否出现过内幕消息。

在给出事件研究的例子之前，我们需要讨论一下**指数**（index number）的概念，以及名义的和实际的经济变量的区别。一个指数一般是浓缩了大量信息的一个数值。在时间序列分析中，特别是在宏观经济应用方面，经常使用指数。指数的一个例子是工业产品指数（IIP），它由联邦储备委员会按月计算。IIP 测量了范围很广的一系列工业的产值，也正是因为这样，它在某一年的大小没有什么数量意义。为了解 IIP 的规模，我们必须知道**基期**

（base period）和**基值**（base value）。在 1997 年总统经济报告中，基准年是 1987 年，基值是 100。（把基期的 IIP 设为 100 只不过是一种习惯；它和把 IIP 设为 1 具有同样的意义，而且确实有一些指数的基值被设为 1。）因为 1992 年的 IIP 是 107.7，我们可以说 1992 年的工业产值比 1987 年高 7.7%。我们可以用 IIP 计算任意两年的工业产值的差别。例如，因为 1970 年的 IIP 是 61.4，1979 年的 IIP 是 85.7，工业生产在（20 世纪）70 年代增长了 39.6%。

改变任何一种指数的基期都很容易。而且，有时我们必须进行这种修改，从而为有不同的基期的指数设定一个统一的基期。例如，如果想把 IIP 的基期从 1987 年改到 1982 年，我们只要把每年的 IIP 的值除以 1982 年的值，然后再乘以 100，使基期值为 100。一般地，公式为

$$newindex_t = 100(oldindex_t / oldindex_{newbase}) \qquad (10.20)$$

328 式中，$oldindex_{newbase}$ 为新基准年的原来的指数值。例如，基年是 1987 年时，1992 年的 IIP 是 107.7，如果我们把基期变为 1982 年，1992 年的 IIP 变成 100(107.7/81.9) = 131.5（因为 1982 年原来的 IIP 为 81.9）。

指数的另一个重要例子是价格指数，如消费价格指数（CPI）。在例 10.1 中我们已经利用 CPI 计算了通货膨胀率。和 IIP 一样，CPI 只有在不同年份或月份（如果使用的是月份数据）之间进行比较时才有意义。在 1997 年的 ERP 中，1970 年的 CPI 是 38.8，1990 年的 CPI 是 130.7。因此，这 20 年间一般价格水平上升了将近 237%。（在 1997 年，CPI 被重新定义，以使 1982 年，1983 年和 1984 年平均值为 100，这样，基期被写成 1982—1984 年。）

除了用来计算通货膨胀率，价格指数在把名义美元数（当前美元数）换算成实际美元数（不变美元数）时也是必须的。大多数经济行为被认为受实际变量而不是名义变量的影响，例如，古典经济学认为劳动供给受实际工资率影响，而不是名义工资率，如果我们有价格指数（如 CPI）的资料，从名义工资率算出实际工资率很容易。首先，要把 CPI 除以 100，以使基期的值为 1，然后，如果 w 代表以名义美元数计算的工资率，由于 $p = \mathrm{CPI}/100$，那么实际工资率就是 w/p，这个工资率是用 CPI 基期的美元数计算的。举个例子，1997 年 ERP 中的表 B—45 中，平均每小时收入用名义金额和 1982 年的美元数（这意味着用来计算实际工资率的 CPI 以 1982 年为基年）分别列了出来。该表显示，1960 年的名义工资率是 2.09 美元/小时，但用 1982 年的美元数衡量则是 6.79 美元/小时。实际工资率在 1973 年达到最高点，用 1982 年的美元数衡量是 8.55 美元/小时，到 1995 年则下降到 7.40 美元/小时。可见，过去 20 年中实际工资率不容忽视地降低了。（如果我们对比一下 1973 年和 1995 年的名义工资，很容易被误导：1973 年为 3.94 美元，1995 年为 11.44 美元。既然实际工资事实上下降了，名义工资的上升就完全由通货膨胀造成。）

标准的经济产出是用实际值表示的。其中最重要的是国内生产总值，即

GDP。常见的出版物上报告的 GDP 的增长一直都是实际 GDP 的增长。1997 年的 ERP 中的表 B—9 把 GDP 用 1992 年的美元数表示出来。在例 10.3 中，我们用同样的方法表示了产量和实际国民生产总值。

实际值变量和自然对数结合起来使用时会发生很有趣的事情，例如，假设平均每周工作小时数与实际工资的关系是

$$\log(hours) = \beta_0 + \beta_1 \log(w/p) + u$$

鉴于$\log(w/p) = \log(w) - \log(p)$，上式可写成

$$\log(hours) = \beta_0 + \beta_1 \log(w) + \beta_2 \log(p) + u \tag{10.21}$$

式中，约束条件是$\beta_2 = -\beta_1$。因此，只有实际工资影响劳动供给的这一假 329 定，对模型（10.21）中的参数施加了限制条件。如果$\beta_2 \neq -\beta_1$，那么价格水平对劳动供给就有影响，这种情形在工人不能区分实际的和名义的工资时就会出现。

具体计算指数的过程中有很多实际问题，这里不作论述。对价格指数的详细讨论在多数中级宏观经济学教材中都找得到，比如 Mankiw（1994，第 2 章）。对我们来说，重要的是要学会在回归分析中使用指数。正如前面提到的，指数的大小不能提供特别多的信息，它们经常以对数形式出现，以便使回归系数可以用百分比变化来解释。

例 10.5　反倾销调查和化学产品进口

克鲁普和波拉德（Krupp and Pollard，1996）分析了美国化学工业对多种进口化学产品的反倾销调查。我们这里把注意力集中在工业用化学产品氯化钡上，它在很多化学过程和石油生产中被用做清洁剂。20 世纪 80 年代，美国的氯化钡生产商认为中国以不合理的低价向美国出口氯化钡（这种行为被称为倾销）。1983 年，他们向美国国际贸易委员会（ITC）提出了反倾销申请。ITC 在 1984 年制定了有利于美国氯化钡生产商的条例。在这个事件中，有一些令人感兴趣的问题，但我们只研究其中的几个。第一，在反倾销调查前的一段时期，进口量不同寻常地高吗？第二，反倾销调查后进口有明显的变化吗？最后，有利于美国的决定执行后，进口究竟减少了多少？

为回答这些问题，我们仿效克鲁普和波拉德定义三个虚拟变量：*befile*6，它在开始调查前的 6 个月为 1；*affile*6 表示开始调查后的 6 个月；*afdec*6 代表调查结束并确认构成倾销行为的 6 个月后。因变量 *chnimp* 是从中国进口的氯化钡的数量，我们用它的对数形式。解释变量包括：化学产量指标 *chempi*，它决定了对氯化钡总的需求；*gas* 代表石油的产出数量，它也是需求变量；汇率指标 *rtwex*，它表示美元对另外几种货币的坚挺程度。1977 年的化学产量指标被定义为 100。我们这里所作的分析与克鲁普和波拉德的区别在于所有的变量都采用自然对数形式（当然，虚拟变量例外），而且，三个虚拟变量都被包括在同一个回归中。

利用 1978 年 2 月到 1988 年 12 月的月度数据，得出

$$
\begin{aligned}
\log(\widehat{chnimp}) = &-17.80 + 3.12\log(chempi) + 0.196\log(gas)\\
&(21.05)\quad (0.48)\qquad\qquad\quad (0.907)\\
&+0.983\log(rtwex) + 0.060befile6 - 0.032affile6\\
&(0.400)\qquad\qquad\quad (0.261)\qquad\quad (0.264)\\
&-0.566afdec6 \qquad\qquad\qquad\qquad\qquad\qquad (10.22)\\
&(0.286)\\
&n = 131, R^2 = 0.305, R^2 = 0.271
\end{aligned}
$$

330 方程显示，*befile*6 在统计上不显著，所以没有证据表明在反倾销调查前的 6 个月来自中国的进口量不同寻常地高。还有，虽然对 *affile*6 的估计为负，但系数非常小（表明从中国的进口下降了 3.2%），它在统计上是不显著的。*afdec*6 的系数显示，作出有利于美国厂商的决定的 6 个月后从中国进口氯化钡的数量大幅度地下降了，这个结果并不令人感到惊奇。既然有这么大效果，我们计算出准确的百分比变化：$100[\exp(-0.566)-1]\approx-43.2\%$。这个系数的双侧检验在 5% 的水平上是显著的。

控制变量系数的符号与我们所预料的一致：总的化学产量的增长增加了对清洁剂的需求；石油产量没有显著地影响从中国的进口；log(*rtwex*)的系数显示，美元相对于其他货币的升值增加了从中国进口的需求，这与经济理论的预测相同。（事实上，弹性并非显著地异于 1。为什么？）

定量变量和定性变量的交互作用也经常在时间序列分析中使用。下面是一个有实际意义的例子。

例 10.6　选举结果和经济形势

费尔（Fair，1996）总结了他在用经济形势来解释总统选举结果方面的工作。他利用 1916—1992 年（每 4 年）的数据得到的 20 次观测，解释了两党选举中民主党获得选票的比例。我们来估计一个简化了的费尔模型（这里使用的变量名比他用的更形象）：

$$
\begin{aligned}
demvote = \beta_0 &+ \beta_1 partyWH + \beta_2 incum + \beta_3 partyWH\cdot gnews\\
&+ \beta_4 partyWH\cdot inf + u
\end{aligned}
$$

式中，*demvote* 为两党选举中民主党获得选票的比例；解释变量 *partyWH* 类似于虚拟变量，但民主党在白宫执政时它的值为 1，共和党执政时其值为 −1。费尔用这个变量规定共和党执政的影响与民主党执政的影响程度相同，但符号相反。从理论上来讲，这是一个很自然的规定，因为两党所得份额之和应为 1。它还节约了两个自由度，这在观测次数如此有限的情况下尤其重要。类似地，变量 *incum* 当执政的民主党参加竞选时定义为 1，当执政的共和党参加竞选时定义为 −1，其他情况时为零。变量 *gnews* 是现任政府执政的前 15 个（总共有 16 个）季度中人均产值增长超过 2.9%（年增长率）的季度数。*inf* 是本届政府的前 15 个季度的年平均通货膨胀率。更准确的定

义参见费尔（Fair，1996）的著作。

经济学家们最感兴趣的是有相互作用的部分 *partyWH*·*gnews* 和 *partyWH*·*inf*。因为 *partyWH* 在民主党在白宫时等于 1，β_3 度量了好的经济消息对执政党的影响；我们预计 $\beta_3>0$。类似地，β_4 度量了通货膨胀对执政党的
331 影响。由于执政期间的通货膨胀被当做坏消息，我们预计 $\beta_4<0$。

用 FAIR.RAW 中的数据估计得出的方程为

$$\begin{aligned}\widehat{demvote}=&\ 0.481-0.043\,5\,partyWH+0.054\,4\,incum\\&(0.012)(0.040\,5)\qquad\quad(0.023\,4)\\&+0.010\,8\,partyWH\cdot gnews-0.007\,7\,partyWH\cdot inf\\&(0.004\,1)\qquad\qquad\qquad(.000\,33)\qquad\qquad(10.23)\\&n=20,R^2=0.663,\bar{R}^2=0.573\end{aligned}$$

除了 *partyWH* 以外，所有变量都在5%的水平上显著。处于执政党的位置可以带来相当于所得选票份额的5.4%的选票。（切记，*demvote* 用得票比例表示。）另外，有价值的经济消息有正的影响：每个季度的好消息相当于1.1个百分点 通货膨胀正如所料，有负的影响：如果平均年通货膨胀上升2个百分点，执政党在选举中会失去1.5个百分点。

我们可以利用这个方程预测1996年在民主党的比尔·克林顿（Bill Clinton）和共和党的鲍伯·多尔（Bob Dole）之间进行的总统选举结果。[自由竞选者罗斯·佩罗特（Ross Perot）被排除了，因为费尔的方程只能用于两党选举的情况。] 既然克林顿以当政者的身份参加竞选，*partyWH* = 1，*incum* = 1。在克林顿执政的前15个季度，人均实际 GDP 有三次超过2.9%，所以 *gnews* = 3。用1997年的 ERP 中的表 B—4 列出的 GDP 缩减指数算出从1991年第四个季度到1996年第三个季度的年均通货膨胀率为3.019（用费尔的公式算出）。把以上代入式（10.23），得

$$\begin{aligned}\widehat{demvote}&=0.481-0.043\,5+0.054\,4+0.010\,8(3)\\&\quad-0.007\,7(3.019)\\&\approx0.501\,1\end{aligned}$$

因此，基于12月份选举前的信息，可以预测到克林顿将获得两党选举中稍微多数的选票：大约50.1%。实际上，他得到了选票的54.65%。

10.5 趋势和季节性

描述有趋势的时间序列

很多经济方面的时间序列沿着时间有上升的一般趋势。为了能用时间序

列数据作出因果性推断，我们必须承认一些时间序列包含有时间趋势。忽略两个序列按相同或相反的趋势发展这一事实会导致错误的结论，认为一个变量的变化由另一个变量的变化引起。在很多情况下，两个时间序列过程表现出相关性，仅仅是因为二者都由于未被观测到的因素的作用而具有某种趋势的缘故。

图 10.2 画出了美国 1947—1987 年的劳动生产率（每小时的产量）的曲线。这个序列呈现出明显的上升趋势，这反映出工人的生产能力随着时间不
332 断提高的事实。

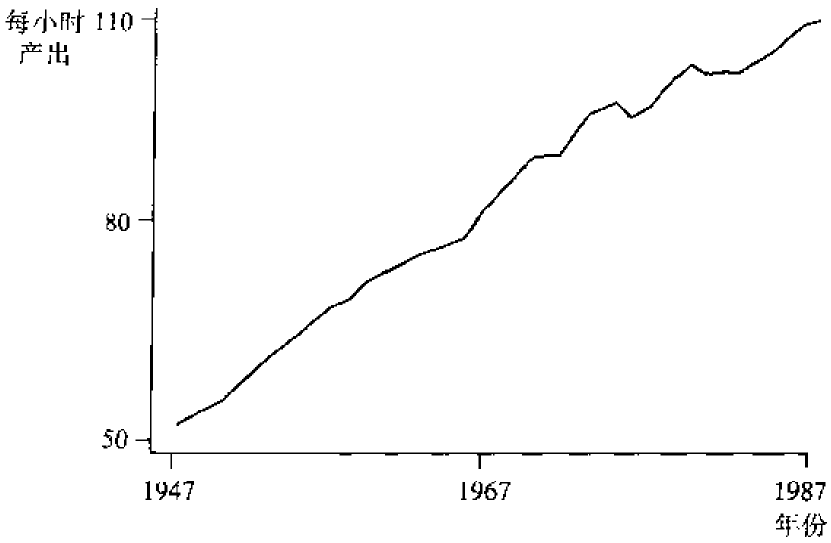


图 10.2　1947—1987 年间美国工人的每小时产出；1977 = 100

也有一些序列，至少在特定时期内，有明显的下降趋势。但是，因为正的趋势更常见，我们将集中研究它们。

什么样的统计模型能够恰当地描述有趋势的行为呢？一个常见的办法是把序列 $\{y_t\}$ 写成

$$y_t = \alpha_0 + \alpha_1 t + e_t, t = 1, 2, \cdots \tag{10.24}$$

在这种最简单的情况下，$\{e_t\}$ 是独立同分布序列（$i.i.d.$），且 $E(e_t) = 0$，$Var(e_t) = \sigma_e^2$。注意参数 α_1 与时间 t 相乘的形式，它形成了一个**线性时间趋势**（linear time trend）。式（10.24）中的 α_1 很容易理解：在其他因素（被包括在 e_t 中）不变时，α_1 度量了由于时间的流逝 y_t 从一个时期到下一个时期的变化：当 $\Delta e_t = 0$ 时，有

$$\Delta y_t = y_t - y_{t-1} = \alpha_1$$

另一种认识具有线性时间趋势的序列的办法是，它的平均值是时间的线性函数：

$$E(y_t) = \alpha_0 + \alpha_1 t \tag{10.25}$$

如果 $\alpha_1 > 0$，那么一般来说，y_t 随着时间增大，所以有向上的趋势。如果 $\alpha_1 < 0$，y_t 就有向下的趋势。y_t 的值因为存在着随机性并不刚好落在式

(10.25) 的线上，但它的期望值落在线上。与平均值不同，y_t 的方差在所有
333 时间上都一样：$\mathrm{Var}(y_t)=\mathrm{Var}(e_t)=\sigma_e^2$。

问题 10.4 在例 10.4 中，我们把总生育率作为一个有限分布滞后模型中的因变量。从 1950 年到 80 年代中期，gfr 有明显的下降趋势。当 $\alpha_1<0$ 时，它在所有将来的时期都会有明显的时间趋势吗？说明理由。

如果 $\{e_t\}$ 是一个 $i.i.d.$ 序列，那么 $\{y_t\}$ 是独立但非同一分布的序列。对有趋势的时间序列的更现实的描述，允许 $\{e_t\}$ 在不同的时间上相关，这种相关并不能改变线性时间趋势的本质。实际上，对经典线性模型假定下的回归分析来说，重要的是 $\mathrm{E}(y_t)$ 与 t 成线性关系。第 11 章学习 OLS 的大样本性质时，我们将讨论 $\{e_t\}$ 容许有多大程度上的时间相关。

很多经济方面的时间序列能用**指数趋势**（exponential trend）来更好地迫近：当一个序列从一个时期到另一个时期的平均增长率为恒定时，它就服从指数趋势。图 10.3 画出了美国 1948—1995 年每年名义进口额（ERP1997，表 B—101）的曲线。

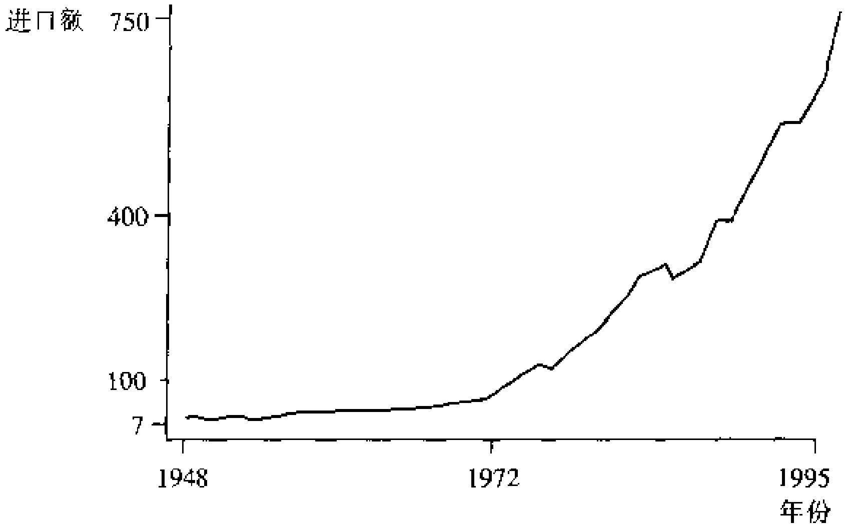


图 10.3 1948—1995 年间美国的名义进口额（以 10 亿美元计）

在较早的年代，每年进口量的变化相对很小，但变化随着时间增大。这与固定平均增长率一致，因为每个时期的变化的百分比大致相同。

在实践中，时间序列中的指数趋势可以通过建立有线性趋势的自然对数模型得到（假设 $y_t>0$）：

334
$$\log(y_t)=\beta_0+\beta_1 t+e_t, \quad t=1,2,\cdots \qquad (10.26)$$

将两边指数化后，y_t 便表现出指数趋势：$y_t=\exp(\beta_0+\beta_1 t+e_t)$。如果我们要在线性回归模型中使用指数趋势时间序列，式（10.26）就是表示这种序列的最方便的办法。

我们怎样解释式（10.26）中的 β_1 呢？对于很小的变化，$\Delta\log(y_t)=$

$\log(y_t)-\log(y_{t-1})$近似等于 y_t 变化的比例：

$$\Delta\log(y_t)\approx(y_t-y_{t-1})/y_{t-1} \tag{10.27}$$

式（10.27）的右边被称做 y 从时期 $t-1$ 到时期 t 的**增长率**（growth rate）。要想把增长率变为百分数，我们只要把它乘以100就行了。如果 y_t 服从（10.26），作此变换并设 $\Delta e_t=0$，得

$$\Delta\log(y_t)=\beta_1 \tag{10.28}$$

也就是说，β_1 近似等于 y_t 各期增长率的平均值。例如，t 代表年份，$\beta_1=0.027$，则 y_t 以平均每年2.7%的速度增长。

尽管线性和指数趋势是最常见的，但实际遇到的趋势可能比它们更加复杂。比如，我们可能会遇到二次时间趋势：

$$y_t=\alpha_0+\alpha_1 t+\alpha_2 t^2+e_t \tag{10.29}$$

如果 α_1 和 α_2 是正的，这个趋势的斜率是增加的，这一点可以通过计算近似斜率（e_t 保持不变）看出：

$$\frac{\Delta y_t}{\Delta t}\approx\alpha_1+2\alpha_2 t \tag{10.30}$$

［如果你对微积分比较熟悉，可以看出式（10.30）的右边是$\alpha_0+\alpha_1 t+\alpha_2 t^2$对 t 的导数。］如果 $\alpha_1>0$，但 $\alpha_2<0$，趋势便呈现隆起的形状。这对某些趋势序列来说可能是不怎么好的描述，因为它要求上升的趋势最终变成下降的趋势。然而，对于有比式（10.24）和式（10.26）更复杂趋势的时间序列来说，在某个特定的时段里，它可能是建立模型的灵活有效的方法。

在回归分析中使用趋势变量

在回归分析中，应该用某种方法把有趋势的被解释或解释变量表示出来，而这并不难办。首先，趋势变量并不一定违背经典线性模型的假定TS.1～TS.6。但是，考虑到可能有一些影响 y 的未被观测到的具有趋势的因素，可能与解释变量相关，我们必须谨慎从事。如果忽略这种可能的情况，我们所找到的 y_t 与一个或多个解释变量之间的关系就会是错误的。仅仅因为每个变量都随着时间增长，而在两个或更多的趋势变量之间找到某种关系的现象是**谬误回归**（spurious regression）的一个例子。所幸的是，只要增加一个时间趋势变量就可以消除这种麻烦。

335 下面举个具体的例子。考虑一个 y_t 受两个可观测变量x_{t1}和x_{t2}影响的模型。而除了这两个变量以外，还存在着随着时间系统地增长或缩减的未被观测到的因素。满足以上特征的模型为

$$y_t=\beta_0+\beta_1 x_{t1}+\beta_2 x_{t2}+\beta_3 t+u_t \tag{10.31}$$

它可以理解成 $x_{t3}=t$ 时的多元线性回归。考虑到方程中的趋势，显然，y_t

可能会基于本质上与 x_{t1} 和 x_{t2} 无关的原因随着时间增长（$\beta_3>0$）或下降（$\beta_3<0$）。如果式（10.31）满足假定 TS.1，TS.2 和 TS.3，那么省略掉 t 只做 y_t 与 x_{t1}，x_{t2} 的回归，一般会得到有偏误的 β_1 和 β_2 的估计值：因为我们实际上已经从回归方程中省略了一个重要的变量 t。这种情况在 x_{t1} 和 x_{t2} 本身有趋势时尤为严重，因为它们可能与 t 高度相关。下面的例子说明了时间趋势如何导致了谬误回归。

例 10.7　房产投资与价格

HSEINV.RAW 中的数据是对美国 1947—1988 年房产投资和房产价格指数的观测结果。令 *invpc* 代表实际人均房产投资（以千美元计），*price* 代表房产价格指数（1982 年为 1）。我们采用一个有固定弹性形式的简单回归方程，它可被看做是一个房产股票的供给方程，得出

$$\begin{aligned}\log(\widehat{invpc}) &= -0.550 + 1.241\log(price)\\ &\quad\ (0.043)\quad\ (0.382)\\ n &= 42, R^2 = 0.208, \bar{R}^2 = 0.189\end{aligned}\tag{10.32}$$

人均投资对价格的弹性非常大，统计上也是显著的；它并非统计上不同于 1。在这里可要小心一些，*invpc* 和 *price* 都有向上的趋势。特别是，如果对 t 做 log(*invpc*) 的回归，得到趋势的系数等于 0.008 1（标准误 = 0.001 8）；log(*price*) 对 t 回归得到趋势系数为 0.004 4（标准误 = 0.000 4）。虽然趋势系数的标准误不一定可靠——这些回归包含了严重的序列相关——但这些系数的估计值的确揭示了向上的趋势。

为了解释变量的趋势化行为，我们增加一个时间趋势：

$$\begin{aligned}\log(\widehat{invpc}) &= -0.913 - 0.381\log(price) + 0.009\,8t\\ &\quad\ (0.136)\quad\ (0.679)\qquad\qquad (0.003\,5)\\ n &= 42, R^2 = 0.341, \bar{R}^2 = 0.307\end{aligned}\tag{10.33}$$

这样一来，结论就不同了：估计出的价格弹性是负的，而且并非显著地异于零；时间趋势是显著的，它的系数表明平均每年 *invpc* 有将近 1% 的增长。从这个分析中，我们无法得出实际人均房产投资受价格影响的结论，因为存在着被包含在时间趋势中的因素，它们能够影响 *invpc*，却没有被包括在模型中。由于价格也随着时间有向上的趋势，式（10.32）的结果展示了 *invpc* 和 *price* 之间的谬误关系。 336

在有些情况下，增加一个趋势可以使关键的解释变量更显著。当自变量和因变量有不相同的趋势（比如一个向上另一个向下），而自变量围绕趋势线的变动会引起因变量向偏离趋势线的方向移动的时候，就会出现这种情况。

例 10.8　生育方程

我们添加一个线性时间趋势到生育方程（10.18）中去，得到

$$\hat{gfr}_t = 111.77 + 0.279\,pe_t - 35.59\,ww2_t + 0.997\,pill_t - 1.15t$$
$$(3.36) \quad (0.040) \quad (6.30) \quad (6.626) \quad (0.19) \qquad (10.34)$$
$$n = 72, R^2 = 0.662, \bar{R}^2 = 0.642$$

上面的 *pe* 的系数是式（10.18）中的估计值的 3 倍多，并且也显著得多。有意思的是，一旦我们加进线性趋势，*pill* 变得不显著了。从估计中可以看出，当其他条件相同时，*gfr* 在这段时期内基本上是不断下降的。

既然总生育率在 1913 年和 1984 年间表现出先上升后下降的趋势，可以采用二次趋势，我们将发现所估计的 *pe* 的影响是那么稳健：

$$\hat{gfr}_t = 124.09 + 0.348\,pe_t - 35.88\,ww2_t - 10.12\,pill_t$$
$$(4.36) \quad (0.040) \quad (5.71) \quad (6.34)$$
$$-2.53t + 0.196t^2$$
$$(0.39) \quad (0.0050) \qquad (10.35)$$
$$n = 72, R^2 = 0.727, \bar{R}^2 = 0.706$$

pe 的系数变得更大了，统计上也更显著了。现在，*pill* 表现出预料中的负影响，并且也达到了显著的边缘（marginally significant），此外，两个趋势项都是显著的。由此可见，二次趋势是一种解释 *gfr* 不平常趋势行为的灵活有效的方法。

你可能会对例 10.8 有疑问：为什么停留在二次趋势的水平上？没有什么能阻止我们添加像 t^3 这样的自变量，实际上，这样做也有可能是合理的（见习题 10.12）。但我们必须得慎重一点，在模型中增加变量时不能太轻率。我们想要得到的是相对简单的趋势，这个趋势应该能够说明未被自变量解释的、较大程度的因变量的变动。如果把足够多的 t 的多次项包括进来，我们可以很好地追踪（track）出任何序列。但这对找出哪个解释变量影响 y_t 没有多大帮助。 337

对有时间趋势的回归做除趋势变换

在回归模型中引进时间趋势，相当于在回归分析中在使用原始数据之前将它们的**趋势**去掉。举个具体的例子：我们虽然只考察式（10.31），但得出的结论具有普遍性。

我们做 y_t 对 x_{t1}，x_{t2}和 t 的回归时，得到拟合方程

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 t \qquad (10.36)$$

我们可以引用第 3 章中对 OLS 偏回归的解释，证明 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 可通过以下步骤得到。

（1）将 y_t，x_{t1}和 x_{t2}分别对常数项和时间趋势 t 做回归，并记录残差 $\ddot{y}_t$，$\ddot{x}_{t1}$和 $\ddot{x}_{t2}$（$t = 1, 2, \cdots, n$）。例如

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t$$

这样，我们可以把 $\ddot{y}_t$ 当成去掉了线性趋势。在去掉 y_t 的趋势的过程中，我们用OLS估计了模型

$$y_t = \alpha_0 + \alpha_1 t + e_t$$

得到的残差 $\hat{e}_t = \ddot{y}_t$ 是去掉了趋势（至少在样本中）的。对 x_{t1} 和 x_{t2} 的解释与此类似。

（2）做回归

$$\ddot{y}_t \text{ 对 } \ddot{x}_{t1} \text{ 和 } \ddot{x}_{t2} \tag{10.37}$$

（截距不是必要的，但是保留截距也没什么影响：估计出来的截距将是零。）这个回归刚好得出式（10.36）中的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$。

这意味着我们最感兴趣的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 可以被理解为来自于一个没有时间趋势的回归，而这个回归中的因变量和自变量都是事先经过了除趋势的。无论有多少个自变量，也无论趋势是二次的还是更高次的，上面这个结论都是成立的。

如果从式（10.36）中漏掉 t，就没有除趋势的过程，y_t 就可能与一个或多个 x_{tj} 相关，造成相关的原因仅仅是因为它们每个都有一个趋势。我们在例10.7中见过这种情况。如果趋势项统计上是显著的，在回归中加入时间趋势会使结果有重大改变，那么在没有考虑到趋势的情况下得到的最初的结果就是令人怀疑的。

$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的推导过程表明，如果自变量含有趋势，即使 y_t 不含趋势，在回归中包括一个时间趋势也不失为一种好办法。假如 y_t 没有明显的趋势，但 x_{t1} 可能随时间上升，那么去掉了回归中的趋势变量，即使沿着趋势变动的 x_t
338 对 y_t 有影响，也会使 x_{t1} 看起来好像对 y_t 没有影响。如果把 t 加到模型中去就不会发生这样的情况。

例 10.9　波多黎各的就业

我们在式（10.17）中加入一个线性趋势，估计结果为

$$\begin{aligned} \log(\widehat{prepop}_t) = &-8.70 - 0.169\log(mincov_t) + 1.06\log(usgnp_t) \\ &\;\;(1.30)\quad\;(0.044)\qquad\qquad\quad\;(0.18) \\ &-0.032t \\ &\;\;(0.005) \\ &n = 38, R^2 = 0.847, \bar{R}^2 = 0.834 \end{aligned} \tag{10.38}$$

log(*usgnp*)的系数有了非常大的变化：从值为 0.012、不显著，到1.06、非常显著。最低工资的系数只有微小的变化，然而标准误明显变小了，而标准误的变小使log(*mincov*)比以前更显著了。

变量 $prepop_t$ 没有表现出明显的向上或向下的趋势，但log(*usgnp*)表现出向上的线性趋势。[做log(*usgnp*)对 t 的回归，得到系数为0.03，因此在这期

间 usgnp 每年增长 3%。] 我们可以这样理解估计值 1.06：当 usgnp 的增长超出长期趋势 1% 时，prepop 增长约 1.06%。

因变量有趋势时**R-**平方的计算

时间序列回归中的 R^2 通常很大，与典型的横截面数据的 R^2 相比尤其大。这意味着我们对时间序列数据中影响 y 的因素知道得更多吗？不一定。一方面，时间序列数据经常是以总量的形式出现（比如美国的平均每小时工资率），而总量通常比个人、家庭或企业的数据容易解释一些，后者常有横截面数据的性质。但是，当因变量有趋势时，时间序列回归中的普通的或校正的 R^2 可能会人为地变大。记住，R^2 是用来度量误差的方差相对于 y 的方差有多大的，这可以通过校正的 R^2 的公式直接看出来：

$$\bar{R}^2 = 1 - (\hat{\sigma}_u^2/\hat{\sigma}_y^2)$$

式中，$\hat{\sigma}_u^2$ 为误差的方差的无偏估计；$\hat{\sigma}_y^2 = \mathrm{SST}/(n-1)$，$\mathrm{SST} = \sum_{t=1}^{n}(y_t - \bar{y})^2$。在 y_t 有趋势时，只要回归中包含了时间趋势，估计误差的方差就不成问题。但是，当 $\mathrm{E}(y_t)$ 有趋势，比如说有线性趋势时［见式（10.24）］，$\mathrm{SST}/(n-1)$ 不再是 $\mathrm{Var}(y_t)$ 的无偏的或一致的估计。事实上，$\mathrm{SST}/(n-1)$ 会严重高估 $\mathrm{Var}(y_t)$，因为它没有考虑到 y_t 的趋势。

339 在因变量有线性、二次或更多阶趋势时，很容易计算出一种“拟合优度”(goodness-of-fit)指标，这种指标事先过滤掉了时间趋势对 y_t 的影响。最简单的一种方法是在一个因变量已经去掉了趋势的回归中计算普通的 R^2。比如，如果模型是式(10.31)，首先就要做 y_t 对 t 的回归，得到残差 $\ddot{y}_t$，然后，回归

$$\ddot{y}_t \text{ 对 } x_{t1},\ x_{t2} \text{ 和 } t \tag{10.39}$$

这个回归的 R^2 是

$$1 - \frac{\mathrm{SSR}}{\sum_{t=1}^{n}\ddot{y}_t^2} \tag{10.40}$$

式中，SSR 等于式（10.36）中的残差的平方和。既然 $\sum_{t=1}^{n}\ddot{y}_t^2 \leqslant \sum_{t=1}^{n}(y_t - \bar{y})^2$（而且经常是严格小于），从式（10.40）得出的 R^2 不大于，而且通常要小于从式（10.36）得出的 R^2。（两个回归中的残差平方和相等。）如果 y_t 有很强的线性时间趋势，式（10.40）会比通常的 R^2 小很多。

式（10.40）中的 R^2 能够更好地反映出 x_{t1} 和 x_{t2} 能在多大程度上解释 y_t，因为它过滤掉了时间趋势的影响。记住，我们总是可以用某种趋势来解释有趋势的变量，但这并不意味着我们发现了引起 y_t 变动的所有因素。我们还可以利用式（10.40）计算出校正的 R^2：把 SSR 除以 $(n-4)$，因为它是

(10.36) 的自由度；把 $\sum_{t=1}^{n} y_t^2$ 除以（$n-2$），因为在除去 y_t 的趋势时要估计两个趋势参数。一般而言，SSR 要除以普通回归（即包括任何一个时间趋势的回归）的自由度，$\sum_{t=1}^{n} y_t^2$ 则除以（$n-p$），这里的 p 是在去掉 y_t 的趋势时使用的趋势参数的个数。关于如何计算有趋势变量的拟合优度指标这一问题的更多讨论，可参见 Wooldridge（1991a）。

例 10.10　房产投资

在例 10.7 中我们看到，在房产投资方程中，连同log(*price*)再加一个时间趋势，对价格弹性就有很大影响。从式（10.33）中得出的 R^2，从字面上告诉我们已经“解释”了log(*invpc*)的变化的 34.1%。但这纯粹是误导。如果我们除掉log(*invpc*)的趋势，将得到的变量对log(*price*)和 t 做回归，R^2 变成了 0.008，而且校正后的 R^2 实际上是负的。因此，log(*price*)围绕它的趋势的变动，对log(*invpc*)围绕它的趋势的变动实际上并没有解释力。这与式（10.33）中log(*price*)的 t 统计量很小这一事实是一致的。

340　在结束这一小节之前，我们必须再说明一点。在计算 F 统计量的 R^2 形式来检验多重假设时，我们只用普通的没有经过除趋势的 R^2。别忘了，使用 R^2 形式的 F 统计量只是计算上的一种策略，因此常用的公式总是合适的。（意思是，F 检验统计量的使用是不成问题的。）

季节性

如果一个时间序列是由定期如每月或每季度（甚至每周或每天）观测而得到的，它就有可能表现出季节性（seasonality）。比如，中西部每月的新房动工（housing start）数量在很大程度上受天气的影响。因为一年中天气的变化（如晴、雨、阴等）没有规律可循（每个月都差不多），我们当然可以认为 1 月份的天气一般要比 6 月份的恶劣，所以 6 月份的房产动工数量一般比一月份高。把这一现象模型化的一种方法是让这个序列 y_t 的期望值随着月份的不同而不同。再举个例子，由于圣诞节的缘故，第四个季度的零售额正常情况下要比前三个季度的高。同样，这种现象也可以通过让平均零售额在一年中随季度而变化反映出来。这种季节平均是在考虑了可能存在的趋势平均之后所要考虑的。例如，最近一年的第一个季度的零售额要比 30 年前的第四个季度的零售额高，这是因为零售额一直在稳步上升。不管怎么样，如果我们对比典型年份中的平均销售额，季节性的节日因素倾向于使第四个季度的零售额变大。

尽管很多月份或季度的数据序列表现出季节性变化，但不是所有的都如此。比如，每月利息率或通货膨胀率并没有明显的季节性变化。而且，有季

节性变化的序列经常在公布之前已经进行了季节性调整。经季节性调整后的一个序列从原理上来讲是已经除掉了季节性因素的。季节性调整可以通过很多方法来完成，更仔细的讨论不在本书研究范围之内。［参见 Harvey（1990）和 Hylleberg（1986）］

季节性调整是如此普遍，以至于在很多情况下我们根本无法获得未经调整的数据。美国的季度 GDP 是个典型的例子，在每年的总统经济报告中，很多宏观经济变量是以每月的频率报告的（至少最近几年是这样的），但具有季节性变化的变量都经过了季节性调整。宏观经济时间序列的主要来源，（包括城市数据库 Citibase），也都对很多序列作季节性调整。因此，需要我们自己去作季节性调整的范围是很有限的。

有时，我们确实要面对一些未经季节调整的数据。知道一些处理回归模型中的季节性的简单方法对我们很有好处。一般来讲，我们可以在模型中包括一组**季节性虚拟变量**（seasonal dummy variables）来解释因变量或自变量中的季节性。

办法很简单。假设我们有月度数据，而且认为每一年中的季节变化大致相同。例如，因为圣诞节总在每年的同一时间到来，我们可以预料一年中的最后几个月的零售额平均来看要比前几个月高。或者，既然不同年份的天气变化非常相似，中西部的新房动工数量平均来看在夏天的几个月要多于冬天的几个月。能够描述这种现象的通用的一个模型为

341

$$y_t = \beta_0 + \delta_1 feb_t + \delta_2 mar_t + \delta_3 apr_t + \cdots + \delta_{11} dec_t + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \tag{10.41}$$

式中，feb_t，mar_t，…，dec_t 为虚拟变量，表示时间 t 是否对应于特定的月份。在公式中，1 月是基准月，β_0 是 1 月份的截距。x_{tj} 确定以后，如果 y_t 没有季节性，从 δ_1 到 δ_{11} 都是零。这很容易通过 F 检验验证。

问题 10.5 ☞ 在方程（10.41）中，3 月份的截距是多少？并解释为什么季节性虚拟变量满足严格外生性假定。

例 10.11　反倾销调查的影响

在例 10.5 中使用了未经季节性调整的月度数据。因此，我们应该添加季节性虚拟变量，看看有没有必要更改结论。比较可能的情况是，平均来看，调查前的几个月比其他月份的进口量不是高了就是低了。我们像式（10.41）中那样添加 11 个月份变量，并检验它们的联合起来的显著性，得到 p 值 = 0.59，所以季节性变量联合起来是不显著的。而且，从统计上的显著性来看，原来的估计中没有什么重要的东西需要改变。克鲁普和波拉德（Krupp and Pollard，1996）用三个变量表示季节（秋季、春季和夏季，冬季作为基准季），而不是用一组月份变量；结果实质上是一样的。

如果数据是按季度给出的，我们必须用三个变量来表示四个季度中的三个，余下的一个作为基准季度。有时，可以允许季节性变量与某些 x_{tj} 有交互作用，从而使 x_{tj} 对 y_t 的影响在一年之内能够变化。

就像在回归中包括一个时间趋势可以理解为首先去掉趋势一样，在回归中加进季节性虚拟变量可以被看做是将数据**除季节性**（deseasonalizing）。举个具体的例子，考察 $k=2$ 时的方程（10.41），x_1 和 x_2 的 OLS 斜率系数 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 能通过以下方法得到：

（1）做 y_t，x_{t2}和 x_{t2}分别对某个常数及月份变量 feb_t，mar_t，…，dec_t 的回归，保留残差为 y_t，x_{t1}和 x_{t2}（$t=1$，2，…，n）。例如

$$y_t = y_t - \hat{a}_0 - \hat{a}_1 feb_t - \hat{a}_2 mar_t - \cdots - \hat{a}_{11} dec_t$$

这是对月份时间序列除季节性的一种方法。对 x_{t1}和 x_{t2}的解释与此相似。

（2）做 y_t 对 x_{t1}和 x_{t2}的回归［与式（10.37）相同］，得到 $\hat{\beta}_1$ 和 $\hat{\beta}_2$。

在有些情况下，如果 y_t 被宣称有季节性，一个更好的“拟合优度”指标是基于除季节性后的 y_t 的 R^2。它过滤掉了不能被 x_{tj}解释的季节性影响。对自由度的调整的论述参见 Wooldridge（1991a）。 342

表现出季节性的时间序列也可能有某种趋势，在这种情况下，我们要用既包含了时间趋势虚拟变量又包括了季节性虚拟变量的回归模型进行估计。这个回归可以转换成使用既被除掉了趋势又被除掉了季节性的序列的回归。伍德里奇（Wooldridge，1991a）讨论了“拟合优度”统计量：实质上是在计算 R^2 之前，我们要对时间趋势和季节性虚拟变量做 y_t 的回归，来达到除趋势和除季节性的目的。

▶ 小 结

在本章，我们学习了使用时间序列数据的基本回归分析。在与横截面分析相似的假定下，OLS 是无偏的（在 TS.1～TS.3 成立时），OLS 是 BLUE 的（在 TS.1～TS.5 成立时），而且，通常的标准误、t 统计量和 F 统计量都可以用于统计推断（在 TS.1～TS.6 成立时）。由于多数时间序列数据中存在着时间上的相关，我们必须对误差与解释变量在所有时期内的关系如何，以及误差本身在时间上的相关关系如何作出假定。在时间序列的应用中，经典线性模型的假定受到很多限制，但它是一个自然的起点。我们把它应用于静态模型，也把它用于有限分布滞后模型中。

对数形式和虚拟变量经常在时间序列分析和事件研究中使用。我们还讨论了指数和用名义的和实际的美元数表示的时间序列。

趋势和季节性在多元回归的框架下是容易处理的，只要在回归方程中加入时间和季节性虚拟变量就行了。我们指出，把普通 R^2 当做“拟合优度”指标存在很多问题；并建议了几种以除趋势和除季节性为基础的方法。

关键术语

自相关
基期
基值
同期外生
除季节性
除趋势
事件研究
指数趋势
有限分布滞后（FDL）模型
增长率
即期乘数
即期倾向
指数
滞后分布
线性时间趋势
长期弹性
长期乘数
长期倾向（LRP）
季节性虚拟变量
季节性
季节性调整
序列相关
短期弹性
谬误回归
静态模型
随机过程
严格外生
时间序列过程
时间趋势

习题

343 10.1 你是否同意以下说法？给出你的判断和简要的说明。

(i) 像横截面观测一样，我们可以假定多数时间序列观测是独立分布的。

(ii) 时间序列回归中的 OLS 估计量在头三个高斯-马尔科夫假定下是无偏的。

(iii) 在多元回归中，一个有趋势的变量不能用做因变量。

(iv) 在使用年度时间序列观测时，不存在季节性的问题。

10.2 令 $gGDP_t$ 表示国内生产总值的年百分比变化，int_t 表示短期利率。假使 $gGDP_t$ 和 int_t 的关系为

$$gGDP_t = \alpha_0 + \delta_0 int_t + \delta_1 int_{t-1} + u_t$$

式中，u_t 与 int_t，int_{t-1} 及其他过去的利率值不相关。再假设美联储依据以下的政策规则：

$$int_t = \gamma_0 + \gamma_1 (gGDP_{t-1} - 3) + v_t, \quad \gamma_1 > 0$$

（当上一年的 GDP 增长率超过 3% 时，联储就会提高利率以防止经济过热。）

如果 v_t 与 int_t 和 u_t 的所有过去值都不相关，请说明 int_t 一定与 u_{t-1} 相关。［提示：将第一个方程作一期滞后，再将得到的 $gGDP_{t-1}$ 代入第二个方程。］这违背了高斯-马尔科夫假定中的哪一个？

10.3　假设 y_t 符合一个二阶 FDL 模型：

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

令 z^* 代表 z_t 的均衡值，y^* 代表 y_t 的均衡值，于是

$$y^* = \alpha_0 + \delta_0 z^* + \delta_1 z^* + \delta_2 z^*$$

证明：由于 z^* 的变化引起的 y^* 的变化等于长期倾向与 z^* 的变化的乘积，即

$$\Delta y^* = \mathrm{LRP} \cdot \Delta z^*$$

它给出了另一种解释 LRP 的方法。

10.4　把三个事件标示变量从方程（10.22）中取出来，我们得到 $R^2 = 0.281$，$R^2 = 0.264$。这些事件标示变量在 10% 的水平上是联合显著的吗？

10.5　假设你有关于新房动工数、利率和实际人均收入的数据。请构造一个可以反映有可能存在于变量中的趋势和季节性的模型。

10.6　在例 10.4 中，我们对分布滞后模型中个别滞后系数的估计是不够准确的。一种消除多重共线性问题的办法是假定 δ_j 服从一种简单的规律。具

344 体来说，考察一个有四期滞后的模型：

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \delta_3 z_{t-3} + \delta_4 z_{t-4} + u_t$$

现在假定 δ_j 是 j 的二次函数：

$$\delta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2$$

γ_0、γ_1 和 γ_2 为参数。这是一个多项式分布滞后（PDL）模型。

（i）将每个 δ_j 的公式代入分布滞后模型中，把它写成用 γ_h（$h = 0,1,2$）表示的模型。

（ii）解释这个用来估计 γ_h 的回归方程。

（iii）上面的多项分布滞后模型是一般模型的一个受约束形式。它受到了多少约束？你如何来检验它们？［提示：用 F 检验。］

计算机习题

10.7　1979 年 10 月联储放弃了以货币供给为目标的政策，开始直接以短期利率为目标。利用 INTDEF.RAW 中的数据，定义一个在 1979 年后的年份里等于 1 的虚拟变量，把这个虚拟变量加到方程（10.15）中去，看一看 1979 年后利率方程有没有什么变化？你能得出什么结论？

10.8　本题利用 BARIUM.RAW 中的数据。

（i）在方程（10.22）中增加一个线性时间趋势。除了趋势变量以外的其他变量是统计上显著的吗？

（ii）在（i）已经估计过的方程中，检验除了趋势变量以外的其他变量的联合

显著性。你能得出什么结论？

(iii)在方程中添加月份虚拟变量，作季节性检验。增加月份虚拟变量对其他估计值及它们的标准误有重要影响吗？

10.9　在最低工资方程(10.38)中加进变量log($prgnp$)。这个变量显著吗？说明它的系数的含义。增加了变量 log($prgnp$)使最低工资的影响有了什么样的改变？

10.10　利用 FERTIL.RAW 中的数据证实方程（10.19）的 LRP 的标准误约为 0.030。

10.11　本题利用 EZANDERS.RAW 中的数据。数据是关于 1980 年 1 月到 1988 年 11 月期间，印第安纳州安德森市每月失业申报数的情况。1984 年，一个工业区（EZ）建立在安德森市（印第安纳州的其他城市也建立了类似的工业区）。[详细情况参见 Papke（1994）。]

（i）对一个时间趋势变量和 11 个月份的虚拟变量做log($uclms$)的回归。这段时期失业申报数的总体趋势如何？（说明时间趋势的系数的含义。）有迹象表明失业申报数有季节性吗？

345 （ii）在（i）的方程中增加一个虚拟变量 ez，并设安德森市有 EZ 的月份的变量值为 1。工业区的存在看起来是否降低了失业申报数？降低了多少？[你应该用第 7 章的公式（7.10）。]

（iii）若要把（ii）中得出的变化全部归功于 EZ 的建立，需要作哪些假定？

10.12　本题利用 FERTIL3.RAW 中的数据。

（i）做 gfr_t 对 t 和 t^2 的回归，并保留残差，得到了去掉趋势的 gfr_t，即 $\ddot{gfr}_t$。

（ii）做 $\ddot{gfr}_t$ 对方程（10.35）中所有变量，包括 t 和 t_2，的回归。比较得出的 R^2 与式（10.35）中的 R^2 有何不同。你有何结论？

（iii）在式（10.35）中加入 t^3 后进行估计。加进的变量统计上显著吗？

10.13　本题利用 CONSUMP.RAW 中的数据组。

（i）估计一个反映实际人均（非耐用品和服务）消费的增长与实际人均可支配收入增长之间关系的简单回归模型，并用对数的变化表示二者。请用标准的格式报告出所得的结果。试解释方程的含义并讨论统计显著性。

（ii）在（i）的方程中添加实际人均可支配收入增长的一期滞后。对消费增长的滞后调整你有何看法？

（iii）在（i）的方程中加入实际利率。你认为它影响消费增长吗？

10.14　本题利用 FERTIL3.RAW 中的数据。

（i）在方程（10.19）中加入 pe_{t-3}和 pe_{t-4}，并检验这些滞后的联合显著性。

（ii）求出（i）中方程的长期倾向及它的标准误。它们与从式（10.19）得出的结果有什么不同？

（iii）估计问题 10.6 中的多项分布滞后模型，求出 LRP 的估计值，并将它与从没有约束条件的模型中得出的结果进行比较。

10.15　本题利用 VOLAT.RAW 中的数据。变量 $rsp500$ 是每月从标准普尔 500 股票指数获得的收益，以年收益率计。(既包括价格变动带来的收益，也包括分得的红利。) 变量 $i3$ 是 3 个月期国债的收益率，$pcip$ 是工业生产的百分比变化，这二者也以年率计。

(i) 考虑方程

$$rsp500_t = \beta_0 + \beta_1 pcip_t + \beta_2 i3_t + u_t$$

你认为 β_1 和 β_2 应该有什么符号?

(ii) 用 OLS 估计上面的方程，用标准的格式报告结果，并解释系数的符号和大小有什么含义。

(iii) 哪些变量是统计上显著的?

346 (iv) 你在（iii）中的发现是否意味着从标准普尔 500 中获得的收益是可预测的? 说明理由。

10.16　利用 INTDEF.RAW 中的数据，考虑模型（10.15)。

(i) 求出这个样本时期内的 inf 和 def 之间的相关程度，并加以评论。

(ii) 在方程中加入 inf 和 def 的一期滞后，再用标准格式报告结果。

(iii) 比较所估计的、代表通货膨胀效应的 LRP 与式（10.15）中相对应的 LRP。二者有很大差别吗?

(iv) 模型中的两期滞后在 5% 的水平上是联合显著的吗?

第 11 章 用时间序列数据计算 OLS 的其他问题

347 在第 10 章，我们讨论了在渐次增强的假定条件下，用于时间序列数据的 OLS 的有限样本性质。在一组完整的时间序列的经典线性模型假定即 TS.1～TS.6 下，OLS 具有与横截面数据条件下同样令人满意的性质。同时，统计推断的方法也与横截面分析中的一样。通过第 5 章的横截面分析，我们知道有很多有利的理由支持我们研究 OLS 的大样本性质。例如，如果误差项不是来自于正态分布，那么我们必须依靠中心极限定理来为常见的 OLS 统计量和置信区间提供依据。

大样本分析在时间序列的情况下显得更加重要。（有讽刺意味的是，大的时间序列样本非常难以得到；但我们除了借助于大样本作近似计算外，没有别的办法。）在 10.3 节，我们解释了在静态和分布滞后模型中严格外生性假定（TS.2）会如何被破坏。我们将在 11.2 节看到，有滞后因变量的模型必定使假定 TS.2 不成立。

不幸的是，时间序列问题的大样本分析比之于横截面问题的大样本分析遇到的困难要多很多。在第 5 章，我们得出了随机抽样条件下 OLS 的大样本性质。当我们容许在不同的时间上的观测值相关时，情况比我们前面研究过的要复杂得多。尽管如此，那些重要的极限定理对某些（并非全部）时间序列过程来说还是成立的。那些定理成立与否的关键在于，不同时期的变量之间的相关性是否足够快地趋于零。在回归分析中对有高度时间相关的时间

序列应给予特别的重视，本章将要提醒读者有关在回归分析中使用这种序列的几个问题

11.1 平稳性和弱相依时间序列

本节中，我们将介绍在用时间序列数据做回归分析时，使用通常的大样本近似所需要的一些重要概念。对这些问题的一般性了解是很重要的，但对其细节不作过高要求。

平稳和非平稳时间序列

318 从历史来看，**平稳过程**（stationary process）的概念在时间序列分析中一直占有重要地位。一个平稳时间序列过程的概率分布具有下述意义的跨时期稳定性：如果我们从序列中任意取出一组随机变量并把这个序列向前移动 h 个时期，其联合概率分布必须保持不变。下面是平稳性的正式定义。

平稳随机过程（stationary stochastic process）：对每一组时间指数$1 \leqslant t_1 < t_2 < \cdots < t_m$，和所有的整数 $h \geqslant 1$，如果$\{x_{t_1}, x_{t_2}, \cdots, x_{t_m}\}$的联合分布与$\{x_{t_1+h}, x_{t_2+h}, \cdots, x_{t_m+h}\}$的联合分布相同，那么随机过程$\{x_t: t=1, 2, \cdots\}$就是平稳的。

这个定义有点抽象，但它的含义很简单。其中一个含义是：（取 $m=1$ 和 $t_1=1$）对所有的 $t=2, 3, \cdots$，x_t 与 x_1 有相同的分布。换句话说，$\{x_t: t=1, 2, \cdots\}$ 序列是同分布的。当然，平稳性的要求不止这些。例如，对任意的 $t \geqslant 1$，(x_1, x_2) 的联合分布（序列中的前两项）必须与 (x_t, x_{t+1}) 的联合分布相同。和前面一样，这里并没有对 x_t 和 x_{t+1} 之间的相关关系作任何限制，而实际上它们可能会高度相关，但平稳性要求所有时期的相邻项之间的相关关系具有相同的性质。

不平稳的随机过程称为**非平稳过程**（nonstationary process）。因为平稳性是潜在的随机过程的而并非可获得的它的某一实现的一个性质，我们很难判断搜集到的数据是否产生于一个平稳的过程。但是，要指出某些序列不是平稳的却很容易。一个具有 10.5 节中谈到的那种类型的时间趋势的过程显然是非平稳的：它的均值随时间而变。

有时，有一种较弱形式的平稳性就足够了。假定$\{x_t: t=1, 2, \cdots\}$的二阶矩是有限的，即$\mathrm{E}(x_t^2)<\infty$对所有 t 成立，那么可以使用下面的定义。

协方差平稳过程：一个有有限二阶矩的随机过程在满足下列条件时是**协方差平稳**（covariance stationary）的：(i) $\mathrm{E}(x_t)$为常数；(ii) $\mathrm{Var}(x_t)$为常数；(iii) 对任何 t，$h \geqslant 1$，$\mathrm{Cov}(x_t, x_{t+h})$只取决于 h，不取决于 t。

协方差平稳只考虑随机过程的前两阶矩：这个过程的均值和方差在不同

的时间都是恒定不变的，而且 x_t 和 x_{t+h} 的协方差只决定于这两项之间的距离 h，与起始时期 t 的位置无关。由此可以推出 x_t 与 x_{t+h} 之间的相关性也只决定于 h。

如果一个平稳过程有有限的二阶矩，那么它一定是协方差平稳的，但反过来未必正确。有时，为强调平稳性是比协方差平稳更强的要求，前者被称为严格平稳（strict stationary）。不过，既然我们不打算深入研究时间序列过程的中心极限定理，就不必顾忌严格平稳和协方差平稳之间的区别：只要一个序列满足任何一种定义，我们都称它是平稳的。

问题 11.1 假使 $\{y_t: t = 1, 2, \cdots\}$ 是由 $y_t = \delta_0 + \delta_1 t + e_t$（$\delta_1 \neq 0$）产生的，$\{e_t: t = 1, 2, \cdots\}$ 是均值为 0，方程为 σ_e^2 的 $i.i.d.$ 序列，那么，(i) $\{y_t\}$ 是协方差平稳的吗？(ii) $y_t - \mathrm{E}(y_t)$ 是协方差平稳的吗？

349 时间序列计量经济学是如何使用平稳性这一概念的呢？在技术层面上，平稳性简化了大数定律和中心极限定理的表述，尽管我们不用对正式的表述担心。在操作层面上，如果我们想通过回归分析掌握两个或更多变量之间的关系，就需要假定某种跨时期的平稳性。如果允许两个变量（比如说 y_t 和 x_t）之间的关系在不同时期随意变化，那么，在只能得到一个单独的时间序列实现的情况下，我们就无法知道一个变量的变化如何影响另一个变量。

实际上，在表述时间序列数据的一个多元回归模型时，我们假定了一定形式的平稳性，即 β_j 不随时间变化。除此以外，假定 TS.4 和 TS.5 意味着误差过程的方差在所有的时间是恒定的；而且两个相邻时期的误差之间的相关程度为零，当然也就不随时间而变。

弱相依时间序列

平稳性涉及在时间移动中过程的联合分布。弱相依则是完全不同的另一个概念。弱相依对在随机变量 x_t 和 x_{t+h} 之间的时间距离 h 变大时这二者之间有多大程度的关系作出限定。对于平稳时间序列来说，弱相依这个概念很容易理解：大概可以理解为，当 h 无限增大时，如果 x_t 和 x_{t+h} 是“近乎独立”的，这个时间序列过程 $\{x_t: t = 1, 2, \cdots\}$ 就被称为**弱相依**（weakly dependent）的。如果序列是非平稳的，作类似“近乎独立”的表述也还是对的。因此，这里我们必须假定近乎独立的概念是与起点 t 无关的。

以上对弱相依的描述有些含糊，但我们无法正式地定义弱相依，因为没有一个定义能包含所有让人感兴趣的方面。有很多特殊形式的弱相依的正式定义，但它们超出了本书的研究范围。［对这些概念更复杂的分析参见 White（1984），Hamilton（1994）和 Wooldridge（1994b）。］

就我们的目的而言，对弱相依的含义有个直观的概念就足够了。可以用相关性来表述协方差平稳序列的特征：如果当 $h \to \infty$ 时 x_t 和 x_{t+h} 之间的相

关性“足够快”地趋于零，协方差平稳时间序列就是弱相依的。（由于协方差平稳的原因，相关性与起点 t 无关。）换句话说，随着变量在时间上的距离变大，相关性变得越来越小。当 $h \to \infty$ 时，$\mathrm{Corr}(x_t, x_{t+h}) \to 0$ 的协方差平稳序列被称为**渐近不相关**（asymptotically uncorrelated）的。这是直观描述弱相依的常见方式。从技术上讲，需要假定相关性足够快地收敛于零，但我们略过不谈。

弱相依为什么对回归分析是重要的呢？从根本上来看，它暗示了大数定律（LLN）和中心极限定理（CLT）的成立，并因此替代了随机抽样的假定。最著名的时间序列数据的中心极限定理，要求平稳性和某种形式的弱相依，因此，平稳的、弱相依的时间序列是多元回归分析中最理想的一种时间序列。在 11.2 节，我们将借助 LLN 和 CLT 证明 OLS 的合理性。不是弱相依的时间序列——在 11.3 节我们将看到这方面的例子——一般不满足 CLT，这就是把它们用于多元回归分析中具有欺骗性的原因。 350

最简单的弱相依时间序列的例子是独立同分布序列：一个独立的序列就是普通的弱相依序列。一个更有意思的弱相依序列的例子是

$$x_t = e_t + \alpha_1 e_{t-1}, \quad t = 1, 2, \cdots \tag{11.1}$$

式中，$\{e_t: t = 0, 1, \cdots\}$ 是均值为 0，方差为 σ_e^2 的独立同分布的序列。过程 $\{x_t\}$ 被称为**一阶移动平均过程**［moving average process of order one，MA(1)］：x_t 是 e_t 和 e_{t-1} 的加权平均值；在下一个时期我们去掉 e_{t-1}，x_{t+1} 决定于 e_{t+1} 和 e_t。把式（11.1）中 e_t 的系数设为 1 并不失一般性。

为什么MA(1)过程是弱相依的呢？序列中的相邻项之间是相关的，这是因为，$x_{t+1} = e_{t+1} + \alpha_1 e_t$，$\mathrm{Cov}(x_t, x_{t+1}) = \alpha_1 \mathrm{Var}(e_t) = \alpha_1 \sigma_e^2$。因为 $\mathrm{Var}(x_t) = (1 + \alpha_1^2)\sigma_e^2$，那么就有 $\mathrm{Corr}(x_t, x_{t-1}) = \alpha_1/(1 + \alpha_1^2)$。例如，如果 $\alpha_1 = 0.5$，就有 $\mathrm{Corr}(x_t, x_{t+1}) = 0.4$。［当 $\alpha_1 = 1$ 时出现最大正相关；此时 $\mathrm{Corr}(x_t, x_{t+1}) = 0.5$。］下面我们看一看序列中距离在两个或两个以上时期的变量：比较明显，它们之间是不相关的，因为它们相互独立。例如，$x_{t+2} = e_{t+2} + \alpha_1 e_{t+1}$ 是独立于 x_t 的，因为 $\{e_t\}$ 是跨时期独立的。由于假定 $\{e_t\}$ 是同分布的，式（11.1）中的 $\{x_t\}$ 显然又是平稳的。因此，MA(1)是平稳的、弱相依的序列，并且可以把大数定律和中心极限定理用于 $\{x_t\}$。

下面的一个过程是更常见的例子：

$$y_t = \rho_1 y_{t-1} + e_t, \quad t = 1, 2, \cdots \tag{11.2}$$

序列的初始点是 $y_0 (t = 0)$，且 $\{e_t: t = 0, 1, \cdots\}$ 是均值为 0，方差为 σ_e^2 的独立同分布的序列。我们还假定 e_t 独立于 y_0 且 $\mathrm{E}(y_0) = 0$。式（11.2）被称为**一阶自回归过程**［autoregressive process of order one，AR(1)］。

如果要AR(1)过程具有弱相依性就需要作一个关键假定，即稳定性条件（stability condition）$|\rho_1| < 1$。一旦假定得到满足，我们称 $\{y_t\}$ 是一个**稳定的AR(1)过程**［stable AR(1) process］。

为了证明稳定的AR(1)过程是渐近不相关的，我们假定它是协方差平稳

的。[实际上，一般来讲 $\{y_t\}$ 是严格平稳的，但证明的技术难度较大。] 我们知道 $\mathrm{E}(y_t)=\mathrm{E}(y_{t-1})$，而在方程（11.2）中 $\rho_1\neq 1$ 只有在 $\mathrm{E}(y_t)=0$ 的情况下才成立。由于 e_t 和 y_{t-1} 相互独立（所以不相关），有 $\mathrm{Var}(y_t)=\rho_1^2\mathrm{Var}(y_{t-1})+\mathrm{Var}(e_t)$。而且，在协方差平稳的条件下，一定有。根据稳定性条件 $\rho_1^2<1$，我们很容易解出 σ_y^2。

351 $$\sigma_y^2=\sigma_e^2/(1-\rho_1^2) \tag{11.3}$$

现在我们可以找出 y_t 和 $y_{t+h}(h\geqslant 1)$ 之间的协方差。反复进行代换，得

$$\begin{aligned}y_{t+h}&=\rho_1 y_{t+h-1}+e_{t+h}=\rho_1(\rho_1 y_{t+h-2}+e_{t+h-1})+e_{t+h}\\&=\rho_1^2 y_{t+h-2}+\rho_1 e_{t+h-1}+e_{t+h}\\&=\cdots\cdots\\&=\rho_1^h y_t+\rho_1^{h-1}e_{t+1}+\cdots+\rho_1 e_{t+h-1}+e_{t+h}\end{aligned}$$

既然对所有 t 有 $\mathrm{E}(y_t)=0$，我们把最后一个式子乘以 y_t 并取期望，求出 $\mathrm{Cov}(y_t,y_{t+h})$。因为 $e_{t+j}(j\geqslant 1)$ 与 y_t 不相关，于是

$$\begin{aligned}\mathrm{Cov}(y_t,y_{t+h})=\mathrm{E}(y_t y_{t+h})&=\rho_1^h\mathrm{E}(y_t^2)+\rho_1^{h-1}\mathrm{E}(y_t e_{t+1})+\cdots\\&+\mathrm{E}(y_t e_{t+h})=\rho_1^h\mathrm{E}(y_t^2)=\rho_1^h\sigma_y^2\end{aligned}$$

既然 σ_y 既是 y_t 又是 y_{t+h} 的标准差，我们可以很容易地找出 $h\geqslant 1$ 时 y_t 和 y_{t+h} 的相关性的大小：

$$\mathrm{Corr}(y_t,y_{t+h})=\mathrm{Cov}(y_t,y_{t+h})/(\sigma_y\sigma_y)=\rho_1^h \tag{11.4}$$

特别地，$\mathrm{Corr}(y_t,y_{t+1})=\rho_1$，所以 ρ_1 是序列中任意两个相邻项的相关系数。

等式（11.4）非常重要，因为它揭示了对 $h\geqslant 1$ 时任意的 y_t 和 y_{t+h} 来说，它们的相关性随着 h 变大而变得非常小：因为 $|\rho_1|<1$，所以当 $h\to\infty$ 时 $\rho_1^h\to 0$，即使 ρ_1 比较大——比如 0.9，它说明相邻项之间有很高的相关性——y_t 和 y_{t+h} 之间的相关性还是会很快地趋于零。举个例子，如果 $\mathrm{Corr}(y_t,y_{t+5})=0.591$，$\mathrm{Corr}(y_t,y_{t+10})=0.349$，$\mathrm{Corr}(y_t,y_{t+20})=0.122$，其中 t 表示年份，前面的式子表明：距离为 20 年的两个年份的产出之间的相关性为 0.122。如果 ρ_1 变得更小，相关性消失得更快。（读者可以试一试 $\rho_1=0.5$ 时的情况来验证这个结论。）

这个分析经验性地证明了稳定的AR(1)过程是弱相依的。AR(1)过程在时间序列数据的多元回归分析中特别重要。我们将在第 12 章介绍它在其他方面的应用，将在第 18 章用它来进行预测。

还有很多其他类型的弱相依时间序列，其中包括自回归和移动平均过程的混合过程。但就我们目前的研究目的而言，前面用到的几种就够了。

在结束本节之前，我们必须强调在时间序列计量经济学中经常引起混淆的一点。趋势序列，虽然必定是非平稳的，但可以是弱相依的。在第 10 章的简单线性时间趋势模型［见方程（10.24）］中，序列 $\{y_t\}$ 就是弱相依的。如果一个序列是弱相依的，而且在除掉了趋势以后是平稳的，我们称它为**趋势—平稳过程**（trend-stationary process）。（这个名字起得不是很形象，

因为我们在假定平稳性的同时也假定了弱相依。）只要模型中加进了适当的趋势，这样的过程就可以像在第10章中那样把它们用于回归分析。

11.2 OLS的渐近性质

352 在第10章我们了解到，在研究某些时间序列问题的时候，经典线性模型的假定无法得到满足。在这种情况下，我们必须像在横截面分析中那样借助于OLS的大样本性质。在本节，我们将阐述几个假定和主要结果，来更一般地证明OLS的合理性。本章中的定理证明起来有些难度，因此我们把它们省略掉，可以参见 Wooldridge（1994b）。

假定TS.1′（线性与弱相依）

除了还需假定$\{(\mathbf{x}_t, y_t): t=1,2,\cdots\}$是弱相依的以外，假定TS.1′和假定TS.1完全相同。换句话说，大数定律和中心极限定理适用于样本均值。

对参数是线性的，这一条件意味着我们可以把模型写成

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \tag{11.5}$$

式中，β_j为需要估计的参数。x_{tj}可以包括滞后自变量和因变量，条件是弱相依假定能得到满足。

我们已经详细讨论过弱相依的概念了，它是一个非常重要的假定。下一节我们将提出几个明显违背了弱相依假定的时间序列过程，并探讨它们在多元回归模型中的应用问题。

假定TS.2′（零条件均值）

对所有的t，$\mathrm{E}(u_t|\mathbf{x}_t)=0$。

这是关于u_t和解释变量之间的关系最自然的假定。它比假定TS.2弱得多，它没有对u_t与其他时期的解释变量之间的关系作任何限制。稍后我们会举个满足假定TS.2′的例子。

下面的一致性结果仅仅要求u_t有零条件均值且与每个x_{tj}不相关——为了某种目的，知道这一含义对我们是有用的：

$$\mathrm{E}(u_t)=0, \mathrm{Cov}(x_{tj}, u_t)=0, \quad j=1,\cdots,k \tag{11.6}$$

我们将主要用到零条件均值假定，在它的基础上可直接进入渐近分析。

假定TS.3′（无完全共线性）

与假定TS.3相同。

353 **定理11.1（OLS的一致性）**

在假定TS.1′，TS.2′和TS.3′成立时，OLS估计量是一致的：$\mathrm{plim}\hat{\beta}_j=\beta_j$（$j=0$，1，…，$k$）。

定理 10.1 和定理 11.1 之间有几个很关键的区别。首先，在定理 11.1 中，我们说 OLS 估计量是一致的，但并不一定是无偏的。再者，在定理 11.1 中，我们没有要求解释变量必须是外生，而是要求潜在的时间序列是弱相依的。当然，弱相依对获得近似的分布结果也是一个比较苛刻的条件，我们将在下面谈到这一点。

例 11.1　静态模型

考虑一个有两个解释变量的静态模型：

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + u_t \tag{11.7}$$

在弱相依条件下，使 OLS 有一致性的充分条件是

$$E(u_t | z_{t1}, z_{t2}) = 0 \tag{11.8}$$

这个条件排除了 u_t 中含有被遗漏且与 z_{t1} 或 z_{t2} 相关的变量的情形。而且 z_{t1} 和 z_{t2} 各自的函数都不会与 u_t 相关，因此，假定 TS.2′排除了错误设定的函数形式，这与横截面的情况相同。其他的问题，如变量 z_{t1} 和 z_{t2} 的测量误差也会导致式（11.8）不成立。

重要的是要知道，假定 TS.2′并没有排除像 u_{t-1} 和 z_{t1} 之间的相关这类情况。如果 z_{t1} 与过去的 y_{t-1} 有关系，这类相关就会出现，例如

$$z_{t1} = \delta_0 + \delta_1 y_{t-1} + v_t \tag{11.9}$$

式中，z_{t1} 可以是一个政策变量，如货币供给的月度百分比变化，而这种变化取决于上个月的通货膨胀率 y_{t-1}。这样的机制一般会导致 z_{t1} 和 u_{t-1} 相关（把 y_{t-1} 替换掉就会发现这一点）。这种反馈在假定 TS.2′下是允许的。

例 11.2　有限分布滞后模型

在有限分布滞后模型

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \tag{11.10}$$

中，一个很自然的假定是，给定当期和所有过去的 z，u_t 的期望值为零：

$$E(u_t | z_t, z_{t-1}, z_{t-2}, z_{t-3}, \cdots) = 0 \tag{11.11}$$

353 这意味着，一旦 z_t，z_{t-1} 和 z_{t-2} 被包括在方程中，没有 z 的其他滞后能够影响 $E(y_t | z_t, z_{t-1}, z_{t-2}, z_{t-3}, \cdots)$；如果不是这样，我们就可以在方程中加入其他的滞后。例如，y_t 为每年投资的百分比变化，z_t 是第 t 年的利率。当我们令 $\mathbf{x}_t = (z_t, z_{t-1}, z_{t-2})$ 时，假定 TS.2′得到满足，于是 OLS 是一致的。正如在前一个例子中一样，TS.2′并不排除 y 对 z 的将来值有反馈作用。

上面的两个例子不一定需要渐近理论作为依据，因为解释变量是严格外

生的。下面一个例子显然违背了严格外生性假定，因此我们只能求助于OLS的大样本性质。

例 11.3　AR(1)模型

考虑下面的AR(1)模型：

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \tag{11.12}$$

式中，在 y 的所有过去值给定时，误差 u_t 的期望值为零：

$$\mathrm{E}(u_t \mid y_{t-1}, y_{t-2}, \cdots) = 0 \tag{11.13}$$

将以上两个等式结合起来，可以看出：

$$\mathrm{E}(y_t \mid y_{t-1}, y_{t-2}, \cdots) = \mathrm{E}(y_t \mid y_{t-1}) = \beta_0 + \beta_1 y_{t-1} \tag{11.14}$$

这个结果非常重要。首先，它意味着一旦 y 的一期滞后得到了控制，就没有 y 的其他滞后能够影响到 y 的期望值。(这就是“一阶”名称的由来。) 第二，关系式被假定为线性的。

既然 $\mathbf{x}_t$ 只包括 y_{t-1}，等式 (11.13) 便意味着假定 TS.2′成立。与此相对照的是，为达到无偏的目的而需要的严格外生性假定即假定 TS.2 却不成
355 立。因为所有时期的解释变量集，除了最后一期的 y 外，就是 $(y_0, y_1, \cdots, y_{n-1})$，所以假定 TS.2 要求对所有的时期 t，u_t 都与 y_0，y_1，…，y_{n-1}之中的任意一个不相关，而AR(1)模型满足不了这个要求。实际上，因为在式 (11.13) 中 u_t 和 y_{t-1}不相关，u_t 和 y_t 就一定相关。基于以上原因，有一期滞后因变量的模型不能满足严格外生性假定 TS.2。

正如11.1节中所述，为了使弱相依条件成立，我们必须假定 $|\beta_1| < 1$。如果这个条件成立，那么定理11.1意味着，通过做 y_t 对 y_{t-1}的回归得到的OLS估计量是 β_0 和 β_1 的一致估计量。不幸的是，$\hat{\beta}_1$ 是有偏误的。而且，如果样本容量 n 比较小或者 β_1 接近于1，偏误就会很大。(对于近似于1的 β_1，$\hat{\beta}_1$ 有严重的向下偏误。) 对中等大小的大样本来说，$\hat{\beta}_1$ 应该是 β_1 的一个较好的估计值。

在使用标准的推断程序时，我们需要作出同方差和无序列相关的假定。这些假定比第10章中的经典线性模型所对应的假定要宽松一些。

假定 TS.4′（同方差）

对所有的 t，有$\mathrm{Var}(u_t \mid \mathbf{x}_t) = \sigma^2$。

假定 TS.5′（无序列相关）

对所有的 $t \neq s$，有$\mathrm{E}(u_t u_s \mid \mathbf{x}_t, \mathbf{x}_s) = 0$。

在假定 TS.4′中，我们只针对时间为 t 的解释变量给出了条件（请与假定 TS.4 对比一下）。在假定 TS.5′中，我们只针对与 u_t 和 u_s 有相同时间的

解释变量给出了条件。这个假定有点难以解释，但是对于研究很多种时间序列回归中 OLS 的大样本性质来说，它是恰当的条件。当我们考虑假定 TS.5′的时候，经常忽略它是以 x_t 和 x_s 为条件的，我们只对所有的 $t \neq s$ 考虑 u_t 和 u_s 是否相关。

序列相关是静态和有限分布滞后回归中经常遇到的麻烦：没有什么能够保证无法观测到的 u_t 在不同时期是不相关的。重要的是，在方程（11.12）和（11.13）表示出的AR(1)模型中，假定 TS.5′成立。既然在时间 t 解释变量是 y_{t-1}，我们必须表明对所有的 $t \neq s$ 都有$\mathrm{E}(u_t u_s | y_{t-1}, y_{s-1}) = 0$。为了证明它，我们假设 $s < t$。（另一种情况是对称的。）因为 $u_s = y_s - \beta_0 - \beta_1 y_{s-1}$，所以 u_s 是在时间 t 以前的 y 的函数。但由式（11.13），$\mathrm{E}(u_t | u_s, y_{t-1}, y_{s-1})$，而重迭期望定律（见附录 B）告诉我们这意味着$\mathrm{E}(u_t u_s | y_{t-1}, y_{s-1}) = 0$。这一点很重要：因为只要式（11.12）中有一期滞后，误差就一定是序列相关的。在 11.4 节我们将更一般化地讨论动态模型的这个特点。

现在，我们得到了一个渐近结果，它与在横截面的情况下实际上是一样的。

定理 11.2（OLS 的渐近正态性）

在假定 TS.1′～TS.5′下，OLS 估计量是渐近正态分布的。而且，通常的 OLS 标准误、t 统计量、F 统计量和 LM 统计量是渐近生效的。

这个定理给第 10 章中至少一部分例子提供了进一步的理论依据：即使经典线性模型假定都不成立，OLS 依然是一致的，通常的推断程序也是生效的。当然，前提是假定 TS.1′～TS.5′都成立。在下一节我们将讨论在哪些情况下弱相依假定可能不成立。序列相关和异方差性的问题将在第 12 章论及。

356 **例 11.4　有效市场假说**

我们可以利用渐近分析来验证一种有效市场假说（efficient market hypothesis, EMH）。令 y_t 为来自于纽约证券市场综合指数的每周百分比收益。有效市场假说的一个静态形式是说，在第 t 周之前的市场上可以观测到的信息应该对预测第 t 周的收益没有任何帮助。如果我们只用过去的 y 的信息，EMH 就可表述为

$$\mathrm{E}(y_t | y_{t-1}, y_{t-2}, \cdots) = \mathrm{E}(y_t) \tag{11.15}$$

如果式（11.15）是错的，我们就可以用过去的每周收益来预测当期的收益。EMH 假定这样的投资机会必定会被发现，并且几乎马上就会消失。

检验式（11.15）的一个简单的方法是，设定式（11.12）中的AR(1)模型作为对立假设。那么，虚拟假设就是 H_0：$\beta_1 = 0$。在虚拟假设下，因为有式（11.15），所以假定 TS.2′成立，于是，正如我们在前面讨论过的那样，不存在序列相关的问题。同方差的假定是$\mathrm{Var}(y_t | y_{t-1}) = \mathrm{Var}(y_t) = \sigma^2$，这里假定它是成立的。在虚拟假设下，股票收益是序列不相关的，所以我们可

以放心地假定它们是弱相依的。根据定理 11.2，我们可以用 $\hat{\beta}_1$ 的 OLS 的 t 统计量来检验 H_0：$\beta_1=0$；H_1：$\beta_1\neq 0$。

NYSE.RAW 中的每周收益是利用从 1976 年 1 月到 1989 年 3 月的数据计算得到的。如果星期三是节假日，在这种少见的情况下，下一个交易日的收盘价就被用来计算收益。在以上一段时期里，平均每周收益为 0.196%，最高的每周收益为 8.45%，最低的是 −15.32%（出现在 1987 年 10 月的股市崩盘期间）。这个AR(1)模型的估计为

$$
\begin{aligned}
\widehat{return}_t &= 0.180 + 0.059\, return_{t-1}\\
&\quad (0.081)\,(0.038) \qquad\qquad (11.16)\\
&\quad n=689,\ R^2=0.003\,5,\ \bar{R}^2=0.002\,0
\end{aligned}
$$

$return_{t-1}$系数的 t 值大约为 1.55，因此，即使在 10%的显著性水平上，在双侧对立假设中 H_0：$\beta_1=0$ 不能被拒绝。从这个估计可以看出，从某一周到下一周的来自纽约证券交易所的股票收益稍微有点正相关，但这种相关性并不是强得使我们有理由拒绝有效市场假说。

在上一个例子中，也许使用AR(1)模型来检验 EMH 的方法不能探测出相隔大于一周的每周收益之间的相关性。其实，用多于一期滞后的模型，估计起来也并不难。例如，一个**二阶自回归模型**（autoregressive model of order two），即AR(2)模型为

$$
\begin{aligned}
&y_t=\beta_0+\beta_1 y_{t-1}+\beta_2 y_{t-2}+u_t\\
&\mathrm{E}(u_t\mid y_{t-1},y_{t-2},\cdots)=0 \qquad\qquad (11.17)
\end{aligned}
$$

357 这里，β_1 和 β_2 还需要满足平稳性条件，这样才能保证AR(2)过程是弱相依的。不过，这并不算什么问题，因为虚拟假设声称 EMH 是成立的：

$$H_0:\beta_1=\beta_2=0 \qquad\qquad (11.18)$$

如果我们再添加一个同方差假定$\mathrm{Var}(u_t\mid y_{t-1},y_{t-2})=\sigma^2$，就可以使用 F 统计量检验式（11.18）了。如果估计一个 $return_t$ 的AR(2)模型，将得到

$$
\begin{aligned}
\widehat{return}_t &= 0.186 + 0.060\, return_{t-1} - 0.038\, return_{t-2}\\
&\quad (0.081)\ (0.038) \qquad\qquad (0.038)\\
&\quad n=688,\ R^2=0.004\,8,\ \bar{R}^2=0.001\,9
\end{aligned}
$$

（这里，因为方程中又多了一期滞后，我们又失去了一次观测）。单独地看，这两期滞后在 10%的水平上都不显著。它们也是联合不显著的：利用 $R^2=0.004\,8$，求出 F 统计量的近似值为 $F=1.65$；这个 F 统计量的 p 值（自由度为 2 和 685）大约是 0.193。因此，即使在 15%的显著性水平上我们都不能拒绝式（11.18）。

例 11.5　附加预期的菲利普斯曲线

一个线性的附加预期的菲利普斯曲线可以写成

$$inf_t - inf_t^e = \beta_1(unem_t - \mu_0) + e_t$$

式中，μ_0 为自然失业率；inf_t^e 为在第 $t-1$ 年形成的预期通货膨胀率。这个模型假定自然失业率是恒定的，宏观经济学家对此持怀疑态度。实际失业和自然失业之间的差距被称为周期性失业（cyclical unemployment）。实际的通货膨胀与预期的通货膨胀之间的差被称为未被预期的通货膨胀（unanticipated inflation）。误差项 e_t 被宏观经济学家称为供给冲击（supply shock）。如果在未被预期到的通货膨胀和周期性失业之间有相互替代关系的话，就有 $\beta_1<0$。[对附加预期的菲利普斯曲线更详尽的讨论参见 Mankiw（1994，11.2 节）。]

为完成这个模型，我们需要对通货膨胀的预期作一个假定。在适应性预期（adaptive expectation）的作用下，当前通货膨胀的预期值决定于最近观测到的通货膨胀。一个极其简单的形式是今年的通货膨胀就是去年的通货膨胀：$inf_t^e = inf_{t-1}$。（在 18.1 节可以看到另一种适应性预期的形成方式。）在以上的假定下，我们可以得到

$$inf_t - inf_{t-1} = \beta_0 + \beta_1 unem_t + e_t$$

或 $$\Delta inf_t = \beta_0 + \beta_1 unem_t + e_t$$

式中，$\Delta inf_t = inf_t - inf_{t-1}, \beta_0 = -\beta_1\mu_0$。（因为 $\beta_1<0$ 而且 $\mu_0>0$，β_0 预期
358 为负。）因此，在适应性预期下，附加预期的菲利普斯曲线揭示了通货膨胀的变化是由于失业水平和供给 e_t 冲击造成的。如果 e_t 与 $unem_t$ 不相关，而通常也是这么假定的，那么我们可以用 OLS 一致地估计出 β_0 和 β_1。（不一定要作其他假定，比如，将来的失业率不受当前供给冲击的影响。）我们假定 TS.1′～TS.5′都成立，那么，估计方程为

$$\begin{aligned}\widehat{\Delta inf}_t &= 3.03 - 0.543 unem_t \\ &\quad(1.38)\quad(0.230) \\ n &= 48,\ R^2 = 0.108,\ \bar{R}^2 = 0.088\end{aligned} \tag{11.19}$$

方程（11.19）揭示出了周期性失业和未被预期的通货膨胀之间的替代关系：*unem* 上升 1 个百分点会使得预期通货膨胀有大于 1.5 个百分点的降幅。这种影响是显著的（双侧 p 值≈0.023）。我们可以把它和例 10.1 中的静态菲利普斯曲线作一下比较，在那里我们发现通货膨胀和失业率之间只有轻微的正向关系。

因为可以把自然失业率写成 $\mu_0 = \beta_0/(-\beta_1)$，我们就可以用式（11.19）来算出自然失业率的估计值：$\hat{\mu}_0 = \hat{\beta}_0/(-\hat{\beta}_1) = 3.03/0.543 \approx 5.58$。我们估计自然失业率大约是 5.6，这正好在宏观经济学家所认可的范围之内：从历史来看，自然失业率一般在 5%～6% 的范围内。我们可以求出这个估计的近似标准误，但计算方法超出了本书的范围。[参见 Davidson 和 MacKinnon（1993）等。]

在假定 TS.1′～TS.5′成立时，我们已经表明了：在定理 5.3 所描述的一堆估计量中，OLS 统计量是渐近有效的，我们只不过把横截面的观测指数 i 换成了时间序列指数 t 罢了。最后，含有有趋势的解释变量的模型也可以满足假定 TS.1′～TS.5′，要求的条件是，它们是趋势平稳的。而且，只要方程中包含了时间趋势，常见的推断程序就是渐近生效的。

问题 11.2 假使期望形成的方式为：$inf_t^e=(1/2)inf_{t-1}+(1/2)inf_{t-2}$，你将用什么样的回归方程来估计附加预期的菲利普斯曲线？

11.3 使用高度持久时间序列做回归分析

前面一节我们阐述了，只要使用的时间序列是弱相依的，常见的 OLS 推断程序在比经典线性模型假定更弱的假定下生效。不幸的是，很多经济的时间序列不是弱相依的。如果第 10 章中的 CLM 假定都成立，在回归分析中使用强相依的时间序列没什么不妥当的。但是，当数据不是弱相依的时候，常见的推断程序对这些假定是否成立非常敏感，因为我们无法借助于大数定律和中心极限定理。在本节，我们举几个**高度持久或强相依**（highly persistent or strongly dependent）时间序列的例子，看看为了用它们做回归分析，我们需要进行怎样的变换。 359

高度持久时间序列

在简单的AR(1)模型（11.2）中，假定 $|\rho_1|<1$ 对于序列是弱相依的是极其关键的。但很多经济的时间序列用 $\rho_1=1$ 的AR(1)模型来描述更好。这时，可以把它们写成

$$y_t=y_{t-1}+e_t,\quad t=1,2,\cdots \tag{11.20}$$

这里，我们再次假定 $\{e_t: t=1,2,\cdots\}$ 是均值为 0、方差为 σ_e^2 的独立同分布序列。我们还假定初始值 y_0 是独立于 e_t（$t\geqslant1$）的。

式（11.20）中的过程被称为一个**随机游走**（random walk）。在这个过程中，时间为 t 的 y 是这样得到的：以前一期的值 y_{t-1} 为基础，加上一个均值为零且独立于 y_{t-1} 的随机变量。随机游走这个名字正是来源于它的这个特征。有时，通过对**新生值**（innovations）e_t 的性质作不同的假定（比如用不相关来代替独立性），可以对随机游走作不同的定义，但是，目前的定义对我们的研究目的来讲已经足够了。

首先，我们求出 y_t 的期望值。利用反复迭代很容易得到

$$y_t = e_t + e_{t-1} + \cdots + e_1 + y_0$$

取两边的期望值，得到

$$\mathrm{E}(y_t) = \mathrm{E}(e_t) + \mathrm{E}(e_{t-1}) + \cdots + \mathrm{E}(e_1) + \mathrm{E}(y_0) = \mathrm{E}(y_0), \quad t \geqslant 1$$

因此，随机游走的期望值不决定于 t。一个常见的假定是 $y_0 = 0$——在时间零过程以零开始——此时，对所有的 t 都有$\mathrm{E}(y_t) = 0$。

然而，随机游走的方差却随着时间而变化。下面我们来计算随机游走的方差。为简单起见，假定 y_0 是非随机的，于是有$\mathrm{Var}(y_0) = 0$；这样做并不会影响到主要结论。那么，根据 $\{e_t\}$ 的 $i.i.d.$ 假定，有

$$\mathrm{Var}(y_t) = \mathrm{Var}(e_t) + \mathrm{Var}(e_{t-1}) + \cdots + \mathrm{Var}(e_1) = \sigma_e^2 t \qquad (11.21)$$

换句话说，随机游走的方差是时间的线性函数，随着时间上升。这表明这个过程不可能是平稳的。

更为重要的是，随机游走表现出了高度持久的行为，这是因为现在的 y 值对于决定很远的将来的 y 值有非常重要的作用。为了说明这一点，我们写出从现在开始以后的 h 期：

$$y_{t+h} = e_{t+h} + e_{t+h-1} + \cdots + e_{t+1} + y_t$$

假定现在时间 t 的当期值是 y_t，我们想要计算出 y_{t+h}的期望值。既然在 y_t 给定时e_{t+j}（$j \geqslant 1$）的期望值是零，我们得到

360

$$\mathrm{E}(y_{t+h} \mid y_t) = y_t, \quad h \geqslant 1 \qquad (11.22)$$

这意味着，无论我们看到多远，y_{t+h}的最好的预测值总是今天的值 y_t。不妨把这种情况与平稳的AR(1)情况相比较，在后面那种情况下，有类似的结论：

$$\mathrm{E}(y_{t+h} \mid y_t) = \rho_1^h y_t, \quad h \geqslant 1$$

在平稳的情况下，$|\rho_1| < 1$，于是，当 $h \to \infty$时$\mathrm{E}(y_t + h \mid y_t)$向零趋近：$y_t$ 的值变得越来越不重要，而且$\mathrm{E}(y_{t+h} \mid y_t)$越来越接近于非条件期望值$\mathrm{E}(y_t) = 0$。

当 $h = 1$ 时，方程（11.22）就与我们在例 11.5 中使用的适应性预期假定一样了：如果通货膨胀符合随机游走，那么，当通货膨胀的所有过去值给定时，inf_t 的期望值就是inf_{t-1}。所以，通货膨胀的随机游走模型为适应性预期提供了依据。

我们还可以看到，当 $\{y_t\}$ 符合随机游走时，对于 t 很大的情形，y_t 和 y_{t-h}的相关性接近于 1。如果$\mathrm{Var}(y_0) = 0$，就会有

$$\mathrm{Corr}(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

所以，它们的相关程度取决于初始点 t［从而 $\{y_t\}$ 不是协方差平稳的］。另外，对于固定的 t，当 $h \to \infty$时，相关性趋于零，但速度不是很快。事实上，t 越大，这种相关性随着 h 变大而趋于零的速度越慢。如果我们让 h 取一个

比较大的值——比如说 $h=100$——我们总可以找到一个足够大的 t，使得 y_t 和 y_{t+h} 之间的相关性任意接近于1。（假使 $h=100$，如果我们要想使得相关性大于0.95的话，只要 $t>1\ 000$ 就可以了。）因此，随机游走不满足渐近不相关序列的要求。

图11.1画出了初始值为 $y_0=0$ 且 $e_t \sim \text{Normal}(0,1)$ 的一个随机游走的两个实现。一般来讲，要想凭对时间序列曲线的观察来判断它是否为随机游走并不是件容易的事。下面，我们将介绍区分弱相依和强相依序列的一种非正式的方法；正式的统计检验将在第18章中介绍。

361

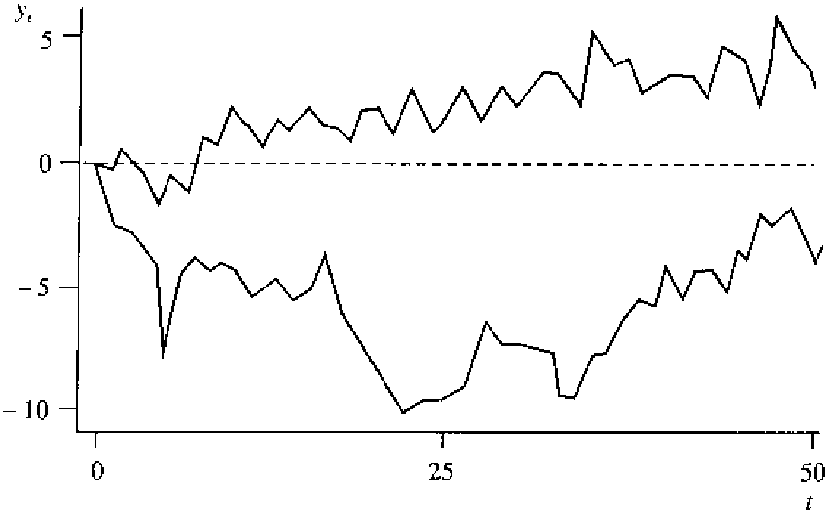


图11.1 随机游走 $y_t=y_{t-1}+e_t$ 的两个实现，$y_0=0, e_t \sim \text{Normal}(0,1), n=50$

公认的、能用随机游走很好地描述的一个序列是3个月期国债利率。图11.2画出了1948—1996年间国债利率的年度数据。

随机游走是所谓的**单位根过程**（unit root process）的一个特例。单位根过程这个名字来源于AR(1)模型中的 $\rho_1=1$ 这一事实。更普通的单位根过程可以像（11.20）中那样产生出来，只不过这里的 $\{e_t\}$ 可以是普通的弱相依序列，而不一定是 *i.i.d.* 序列。[例如，$\{e_t\}$ 可以遵循MA(1)或平稳的AR(1)过程。] 当 $\{e_t\}$ 不是一个 *i.i.d.* 序列的时候，我们在前面得到的随机游走的性质不再成立。但 $\{y_t\}$ 的关键特征能得以保留：现在的 y 值与即使是很远的将来的 y 都高度相关。

从政策角度来看，知道一个经济的时间序列是否高度持久往往是很重要的。我们就以美国的国内生产总值为例。如果GDP是渐近不相关的，那么下一年的GDP水平顶多与很多年之前（例如30年前）的GDP有弱相依的关系。这意味着一项很久以前作用了GDP的政策几乎没有什么持续的影响。相反，如果GDP是强相依的，那么下一年的GDP可以与很多年前的GDP高度相关。正是基于以上原因，我们有必要判断一项造成GDP跳跃变化的政策是否具有长期影响。

非常重要的是，千万不能混淆有趋势的行为和高度持久的行为。有趋势

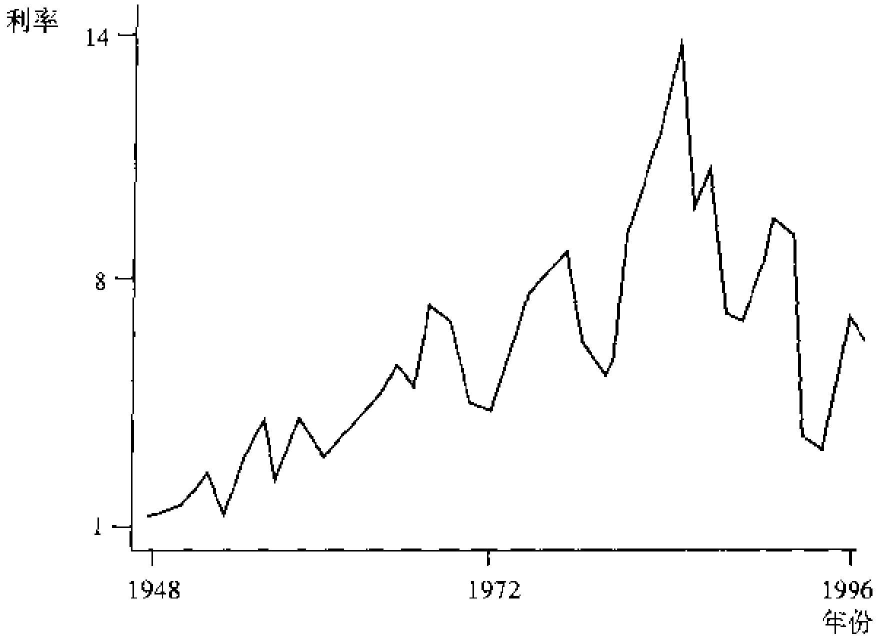


图 11.2　美国 3 个月国债利率，1948—1996 年

的序列不一定是高度持久的，这一点我们在第 10 章中已经了解了。同时，很多人认为像利率、通货膨胀率和失业率这些因素是高度持久的，但是，它们又没有明显的向上或向下的趋势。然而，更为常见的情况是高度持久的序列包含了明显的趋势。**有漂移的随机游走**（random walk with drift）是一个导致了这种行为的模型：

$$y_t = \alpha_0 + y_{t-1} + e_t, \qquad t = 1, 2, \cdots \tag{11.23}$$

式中，$\{e_t: t = 1, 2, \cdots\}$，而 y_0 具有与随机游走模型相同的性质。参数 α_0 是个新东西，它被称做漂移项（drift term）。也就是说，为了得到 y_t，乘数常数 α_0 是连同随机噪声一道加到前一期的值 y_{t-1} 上的。通过反复迭代，我们发现 y_t 的期望值具有一种线性时间趋势：

$$y_t = \alpha_0 t + e_t + e_{t-1} + \cdots + e_1 + y_0$$

因此，如果 $y_0 = 0$，就有 $E(y_t) = \alpha_0 t$，在这种条件下：如果 $\alpha_0 > 0$，y_t 的期望值随时间增长；如果 $\alpha_0 < 0$，它又随时间下降。利用在单纯随机游走情况下的推算方法，我们可以得到 $E(y_{t+h} | y_t) = \alpha_0 h + y_t$，所以 y_{t+h} 在时间 t 的最佳的预测值是 y_t 再加上漂移 $\alpha_0 h$。y_t 的方差与纯粹随机游走情况下的方差完全相同。

图 11.3 描述了有漂移的随机游走的一个实现，其中，$n = 50$，$y_0 = 0$，$\alpha_0 = 2$，e_t 是遵循 Normal（0，9）的随机变量。正如我们从图中看到的，y_t 倾向于随时间增长，但这些序列并不有规律地向趋势线靠拢。

有漂移的随机游走是单位根过程的另一个例子，因为它是在 AR(1) 模型中 $\rho_1 = 1$ 且有截距的特例：

$$y_t = \alpha_0 + \rho_1 y_{t-1} + e_t$$

当 $\rho_1 = 1$ 而 $\{e_t\}$ 是任意的弱相依过程时，我们得到了一大类均值有线性趋
363 势的、高度持久的时间序列过程。

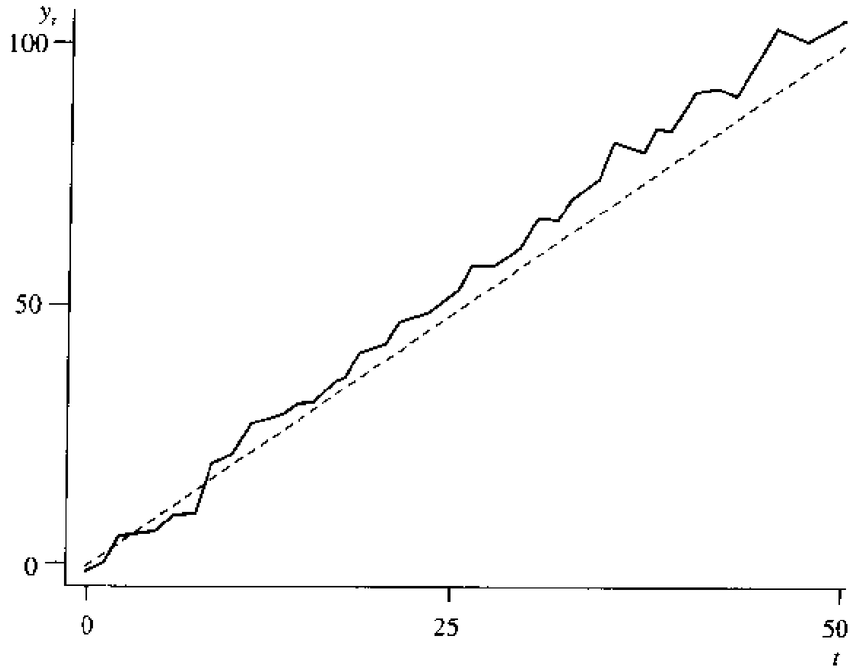


图 11.3　有漂移的随机游走的一个实现，$y_t = 2 + y_{t-1} + e_t, y_0 = 0$，$e_t \sim \text{Normal}(0,9), n = 50$；虚线是 y_t 的期望值，$\mathrm{E}(y_t) = 2t$

高度持久时间序列的变换

可以用单位根过程表示的一类高度持久的时间序列，在不满足 CLM 假定的情况下，一旦用于回归方程中，可能导致误导性结果。在第 18 章我们将更详细地研究谬误回归的问题，现在只要知道可能存在的潜在问题就行了。令人高兴的是，只要作一些简单的变换，就可以使单位根过程变为弱相依的了。

弱相依过程被称为**零阶积整** [integrated of order zero, I(0)]。它的意思是，在回归分析中使用之前无须对这种序列进行任何处理：这种序列的均值已经满足标准的极限定理。单位根过程，例如随机游走（有或没有漂移）被称做**一阶积整** [integrated of order zero, I(1)]，这意味着这个过程的**一阶差分**（first difference）是弱相依的（而且通常是平稳的）。

对于一个随机游走来说，这一点很容易就可以看出来。如果 $\{y_t\}$ 像在式（11.20）中那样产生，就有

$$\Delta y_t = y_t - y_{t-1} = e_t, \quad t = 2, 3, \cdots \tag{11.24}$$

因此，一阶差分后的序列 $\{\Delta y_t : t = 2, 3, \cdots\}$ 实际上是 *i.i.d.* 序列。更一

般地讲，如果 $\{y_t\}$ 是由式（11.24）产生的，其中 $\{e_t\}$ 是弱相依过程，那么，$\{\Delta y_t\}$ 就是弱相依的。因此，当我们怀疑某些过程是一阶积整时，我们经常进行一阶差分，从而使它们可以用于回归分析。稍后，我们将看几个例子。

很多严格正的时间序列 y_t 是这样的：$\log(y_t)$ 是一阶积整。在这种情况下，我们可以在回归分析中使用对数的差分，$\Delta\log(y_t)=\log(y_t)-\log(y_{t-1})$。同时，因为

$$\Delta\log(y_t)\approx(y_t-y_{t-1})/y_{t-1} \tag{11.25}$$

364 可以直接使用 y_t 的变化比例或百分比；在例 11.4 中就是这么做的，在那里我们并没有用股票价格 p_t 来表述有效市场假设，而是用了每周价格变化的百分比 $return_t=100[(p_t-p_{t-1})/p_{t-1}]$。

在回归分析中使用时间序列之前就对它们进行差分还有另一个好处：它除掉了所有的线性时间趋势。这一点很容易通过把有线性趋势的变量写成下面的形式看出来：

$$y_t=\gamma_0+\gamma_1 t+v_t$$

式中，v_t 的均值为零。于是，$\Delta y_t=\gamma_1+\Delta v_t$，所以 $\mathrm{E}(\Delta y_t)=\gamma_1+\mathrm{E}(\Delta v_t)=\gamma_1$。换句话说，$\mathrm{E}(\Delta y_t)$ 是恒定不变的。当 $\Delta\log(y_t)$ 符合线性时间趋势时，对 $\Delta\log(y_t)$ 有同样的结论。因此，我们可以不在回归中引进时间趋势，只要对那些表现出明显的趋势的变量进行差分就行了。

判断时间序列是否是I(1)

判断某个特定的时间序列的实现是I(1)的结果还是I(0)的结果非常困难。统计检验可以用来解决这一问题，但有较大难度；我们将在第 18 章介绍一种简单的处理方法。

还有一些不标准的方法，可以提供实用的判断原则，以确定一个时间序列是否可以用弱相依来大致描述。一个非常简单的工具是受AR(1)模型启发得来的：如果 $|\rho_1|<1$，那么这个过程是I(0)；如果 $\rho_1=1$，那么它是I(1)。我们在前面已经说明了，当AR(1)过程是平稳的时候，$\rho_1=\mathrm{Corr}(y_t,y_{t-1})$。因此，我们可以通过 y_t 和 y_{t-1} 的样本相关性来估计 ρ_1。这个样本相关系数被称为 $\{y_t\}$ 的**一阶自相关**（first order autocorrelation），我们用 $\hat{\rho}_1$ 来表示它。通过应用大数定律，可以证明当 $|\rho_1|<1$ 时，$\hat{\rho}_1$ 是 ρ_1 的一致估计。（但 $\hat{\rho}_1$ 不是 ρ_1 的无偏估计量。）

我们可以借助 $\hat{\rho}_1$ 的值来判断一个过程是I(1)还是I(0)。不幸的是，因为 $\hat{\rho}_1$ 是一个估计量，我们永远也无法有把握地断定是否有 $\rho_1<1$。比较理想的情况是，我们可以计算 ρ_1 的置信区间，看看它是否包含 $\rho_1=1$ 这个值，但这么做很困难：当 ρ_1 接近于 1 和 ρ_1 比 1 小很多时，估计量 $\hat{\rho}_1$ 的样本分布

有非常大的不同。（实际上，当 ρ_1 接近于 1 时 $\hat{\rho}_1$ 可能有一种严重的向下偏误。）

在第 18 章，我们将说明如何检验 H_0：$\rho_1 = 1$ 对立于 H_1：$\rho_1 < 1$。目前，我们只能把 $\hat{\rho}_1$ 当做一个大致的标准，用它来判断某个序列是否需要差分。在作出这样的判断方面并没有很明确的规则存在。大多数经济学家认为，如果 $\hat{\rho}_1 > 0.9$，就需要进行差分；也有些人认为只要 $\hat{\rho}_1 > 0.8$ 就应该差分。

例 11.6　生育方程

在例 10.4 中，我们用个人豁免金额 *pe* 解释了总生育率 *gfr*。这些序列的一阶自相关程度都很高：*gfr* 的 $\hat{\rho}_1 = 0.977$，*pe* 的 $\hat{\rho}_1 = 0.964$。这些都暗示了单位根性态的存在，它使我们对第 10 章中使用的常见的 OLS 的 t 统计量产生怀疑。我们现在用一阶差分的方法来估计这个方程（为了简单起见，我们省略掉了虚拟变量）：

365

$$\begin{aligned} \Delta \widehat{gfr} &= -0.785 - 0.043\Delta pe \\ &\quad\ (0.502)\quad (0.028) \\ &\quad\ n = 71, R^2 = 0.032, \bar{R}^2 = 0.018 \end{aligned} \tag{11.26}$$

根据这个估计，*pe* 的增加会降低同期的 *gfr*，尽管这个估计在 5% 的水平上不是统计上异于零的。这个结果，与我们以前在水平形式上估计模型而得出的结果有很大的不同，它使我们对以前的分析产生怀疑。

如果我们加进 Δpe 的两期滞后，情况就会好些：

$$\begin{aligned} \Delta \widehat{gfr} &= -0.964 - 0.036\Delta pe - 0.014\Delta pe_{-1} + 0.110\Delta pe_{-2} \\ &\quad\ (0.468)\quad (.027)\qquad (.028)\qquad\quad (.027) \\ &\quad\ n = 69, R^2 = 0.233, \bar{R}^2 = 0.197 \end{aligned} \tag{11.27}$$

尽管 Δpe 和 Δpe_{-1} 有负的系数，但它们的系数很小，而且是联合不显著的（p 值 $= 0.28$）。第二期滞后非常显著，表明了 *pe* 的变化与两年后的 *gfr* 的变化之间的正向关系。这比当期就看到影响更为合理。对这个方程的一阶差分形式的分析见习题 11.12。

当我们研究的序列有明显的向上或向下的趋势时，在经过除趋势后得到的一阶自相关更有价值。如果数据没有经过除趋势，自回归相关性倾向于被高估，这会带来偏差，使得我们在有趋势的过程中找到单位根。

例 11.7　工资和生产率

变量 *hrwage* 是美国的平均每小时工资，*outphr* 是每小时产出。估计每小时工资对每小时产出的弹性的一个办法是估计下面的方程：

$$\log(hrwage_t) = \beta_0 + \beta_1 \log(outphr_t) + \beta_2 t + u_t$$

这个方程包含了时间趋势，因为 $\log(hrwage)$ 和 $\log(outphr_t)$ 都表现出明

显的向上的线性趋势。利用 EARNS.RAW 中 1947—1987 年的数据，我们得到

$$\begin{aligned}\log(hrwage_t) &= -5.33 + 1.64\log(outphr_t) - 0.018t \\ &\quad\ (0.37)\quad (0.09) \qquad\qquad\quad (0.002) \qquad\qquad (11.28) \\ n &= 41, R^2 = 0.971, \bar{R}^2 = 0.970\end{aligned}$$

（我们在这里报告了常见的拟合优度指标；如果能像在 10.5 节那样，在对变量除趋势基础上报告这些指标，那就更好了。）估计得到的弹性看起来太大了：生产率增长 1% 使实际工资增长 1.64%。因为标准误很小，以至于95% 366 置信区间很容易就排除了单位弹性的可能。也许是因为美国的工人不相信生产率每增长 1% 他们的工资就上升 1.5% 以上。

要慎重看待式（11.28）中的回归结果。即使对$\log(hrwage)$除线性趋势后，一阶自相关仍为 0.967；而且，对于除掉了趋势的$\log(outphr)$，$\hat{\rho}_1 = 0.954$。这些都表明这两个序列都有单位根，所以我们用一阶差分来重新估计这个方程（我们不再需要时间趋势了）：

$$\begin{aligned}\Delta\log(hrwage_t) &= -0.003\,6 + 0.809\Delta\log(outphr) \\ &\quad\ (0.004\,2) \quad (0.173) \qquad\qquad\qquad (11.29) \\ n &= 40, R^2 = 0.364, \bar{R}^2 = 0.348\end{aligned}$$

根据估计，生产率增长 1% 会使实际工资上升 0.81%，而且估计值不是统计上异于 1。校正后的 R^2 说明，产出的增长能够解释实际工资上升的 35%。在习题 11.9 中，你可以看到一个简单的使用了一阶差分的分布滞后模型。

在前面两个例子中，无论是因变量还是自变量都有单位根。在其他情况下，我们可能会碰到既有单位根又是弱相依的（尽管可能有趋势）的混合型的过程，习题 11.8 就是这方面的一个例子。

11.4 动态完整模型和无序列相关

在AR(1)模型（11.12）中，我们表明了：在假定（11.13）下，误差$\{u_t\}$一定是**序列不相关的**（serially uncorrelated），也就是假定 TS.5′得到满足——假定不存在序列相关。它实际上与假定 E（$y_t | y_{t-1}, y_{t-2}, \cdots$）中只包含 y 的一期滞后是一回事。

我们可以对其他回归模型作类似的表述吗？答案是肯定的。考虑下面简单的静态模型：

$$y_t = \beta_0 + \beta_1 z_t + u_t \qquad (11.30)$$

式中，y_t 和 z_t 作同期标注。若要 OLS 是一致的，只须$E(u_t | z_t) = 0$。通常

情况下，$\{u_t\}$ 是序列相关的。当时，我们在这里就假定：

$$E(u_t \mid z_t, y_{t-1}, z_{t-1}, \cdots) = 0 \tag{11.31}$$

那么，（正如我们在后面一般性地说明的那样）假定 TS.5′成立。尤其是，$\{u_t\}$ 是序列不相关的。

为了研究式（11.31）的含义，我们将式（11.30）和式（11.31）等价地表示为

$$E(y_t \mid z_t, y_{t-1}, z_{t-1}, \cdots) = E(y_t \mid z_t) = \beta_0 + \beta_1 z_t \tag{11.32}$$

367 其中，第一个等式是我们现在感兴趣的问题。它的含义是，一旦 z_t 被确定下来，没有 y 或 z 的滞后能够解释当前的 y。这个要求实际上很严格；如果它不成立，误差就会是序列相关的。

下面，考虑一个有两期滞后的有限分布滞后模型：

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t \tag{11.33}$$

既然我们希望找到 z 对 y 的滞后影响，就很自然地假定式（11.33）描述了分布滞后动态（distributed lag dynamics）：

$$E(y_t \mid z_t, z_{t-1}, z_{t-2}, z_{t-3}, \cdots) = E(y_t \mid z_t, z_{t-1}, z_{t-2}) \tag{11.34}$$

就是说，最多 z 有两期滞后对 y 有影响。如果式（11.31）成立，我们还可以作进一步的表述：一旦我们把 z 和它的两期滞后确定下来，y 的滞后和 z 的其他滞后都无法影响当前的 y：

$$E(y_t \mid z_t, y_{t-1}, z_{t-1}, \cdots) = E(y_t \mid z_t, z_{t-1}, z_{t-2}) \tag{11.35}$$

方程（11.35）比方程（11.32）可能性更大些，但它同样排除了 y 的滞后对当前的 y 的影响。接下来，考虑一个分别有 y 和 z 的一期滞后的模型：

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t$$

既然这个模型中只包括一期滞后变量，式（11.31）是个很自然的假定，也就意味着

$$E(y_t \mid z_t, y_{t-1}, z_{t-1}, y_{t-2}, \cdots) = E(y_t \mid z_t, y_{t-1}, z_{t-1})$$

换句话说，一旦 z_t，y_{t-1}和 z_{t-1}被确定下来，没有 y 或 z 的其他滞后能够影响当前的 y。

在一般化的模型中

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \tag{11.36}$$

这里的解释变量 $\mathbf{x}_t = (x_{t1}, \cdots, x_{tk})$ 可能包含也可能不包含 y 或 z 的滞后，式（11.31）就变成了

$$E(u_t \mid \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \cdots) = 0 \tag{11.37}$$

用 y_t 表示为

$$E(y_t \mid \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \cdots) = E(y_t \mid \mathbf{x}_t) \tag{11.38}$$

368 也就是说，无论 x_t 包含了什么，它都包括了足够多的滞后，以至于 y 和解释变量的其他滞后对解释 y_t 都没有任何意义。当这个条件成立时，我们得到了一个**动态完整模型**（dynamically complete model）。正如我们前面看到的，动态完整性对静态模型和有限分布滞后模型来说是非常强的假定。

一旦我们开始把滞后的 y 当成解释变量，通常认为这个模型是动态完整的。我们将在第18章谈到这个规则的例外情况。

因为式（11.37）等价于

$$E(u_t|x_t,u_{t-1},x_{t-1},u_{t-2},\cdots)=0 \tag{11.39}$$

我们可以证明动态完整模型一定满足假定TS.5′。（它的推导并不重要，可以将它跳过去。）具体来看，设 $s<t$，利用重期望定律（见附录B），得出

$$\begin{aligned}E(u_tu_s|x_t,x_s)&=E[E(u_tu_s|x_t,x_s,u_s)|x_t,x_s]\\&=E[u_sE(u_t|x_t,x_s,u_s)|x_t,x_s]\end{aligned}$$

其中第二个等式可由 $E(u_tu_s|x_t,x_s,u_s)=u_sE(u_t|x_t,x_s,u_s)$ 推出。既然 $s<t$，$(x_t,\ x_s,\ u_s)$ 是式（11.39）中的条件集（conditioning set）的子集。因此，式（11.39）暗示了 $E(u_t|x_t,x_s,u_s)=0$，于是

$$E(u_tu_s|x_t,x_s)=E(u_s\cdot 0|x_t,x_s)=0$$

它说明假定TS.5′成立。

问题 11.3

如果在式（11.33）中，$u_t=e_t+\alpha_1e_{t-1}$，而 $\{e_t\}$ 是均值为0，方差为 σ_e^2 的 *i.i.d.* 序列，那么，方程（11.33）是动态完整的吗？

既然设定一个动态完整模型意味着不存在序列相关，那么是否所有的模型都应该是动态完整的呢？正如我们在第18章将要看到的，对于预测的目的而言，答案是肯定的。一些人认为所有的模型都应该是动态完整的，而一个模型中误差的序列相关是错误设定的标志。但这种观点过于苛刻，有时，我们只对静态模型（例如菲利普斯曲线）或有限分布滞后模型（例如计算生产率上升1%所导致的工资长期变化的百分比）感兴趣。在下一章，我们将讨论如何探测和纠正这些模型中的序列相关。

例 11.8 生育方程

在方程（11.27）中，我们估计了一个 Δgfr 对 Δpe 的分布滞后模型，并包含了 Δpe 的两期滞后。为了让这个模型在式（11.38）的意义上是动态完整的，Δgfr 的滞后和 Δpe 的其他滞后都不应出现在模型中。我们很容易通过添加 Δgfr_{-1} 看出这个结论是错的：添加的滞后的系数的估计值是0.300，t 统计量是2.84。所以，从式（11.38）的意义上来看，这个模型不是动态完整的。

我们能从这里学到些什么呢？对有滞后因变量的一般模型的解释，将放

在第 18 章来讲。就本例来看，式（11.27）不是动态完整的这一事实，暗示了误差可能存在着序列相关。我们将在第 12 章讲述如何检验和纠正这类问题。

11.5 时间序列模型的同方差假定

369 时间序列回归中的同方差假定，特别是 TS.4′，看起来与横截面回归中的同方差假定非常相似。但是，x_t 可能既包括滞后的 y，又包括滞后的解释变量，我们简要地讨论一下在不同时间序列回归中同方差假定的含义。

在下面的简单的静态模型

$$y_t=\beta_0+\beta_1 z_t+u_t \tag{11.40}$$

中，假定 TS.4′要求

$$\mathrm{Var}(u_t|z_t)=\sigma^2$$

因此，即使E$(y_t|z_t)$是 z_t 的线性函数，Var$(y_t|z_t)$一定要是恒定不变的。这一点很容易理解。

在例 11.4 中，我们谈到，对于AR(1)模型（11.12），同方差假定是

$$\mathrm{Var}(u_t|y_{t-1})=\mathrm{Var}(y_t|y_{t-1})=\sigma^2$$

即使E$(y_t|y_{t-1})$取决于 y_{t-1}，Var$(y_t|y_{t-1})$也不取决于 y_{t-1}。所以，y_t 的分布的变异必定与y_{t-1}无关。

现在应该看得清楚，如果我们有模型

$$y_t=\beta_0+\beta_1 z_t+\beta_2 y_{t-1}+\beta_3 z_{t-1}+u_t$$

相应的同方差假定就是

$$\mathrm{Var}(u_t|z_t,y_{t-1},z_{t-1})=\mathrm{Var}(y_t|z_t,y_{t-1},z_{t-1})=\sigma^2$$

它使得 u_t 的方差不决定于z_t，y_{t-1}和 z_{t-1}（或时间的其他函数）。一般情况下，无论模型中出现什么样的解释变量，我们必须假定在解释变量给定时 y_t 的方差是个常数。如果模型中包含滞后的 y 或滞后的解释变量，我们就需要明确地排除异方差的动态形式（这是我们将在第 12 章学习的内容）。但是，在静态模型中，我们只关心Var$(y_t|z_t)$。在方程（11.37）中，我们并不需要对诸如Var$(y_t|y_{t-1})$之类的东西直接作出任何限制。

▶ 小结

本章阐明了，在某些条件能得到满足的情况下，利用渐近分析可以证明

OLS的合理性。理想的时间序列过程是平稳和弱相依的，但是平稳性并不是很关键，要想应用标准的大样本结果，特别是中心极限定理，弱相依是必须的。

如果有趋势的过程是弱相依的，只要模型中包含了时间趋势（像10.5节中那样），这些过程就可以直接用于回归分析中。对有季节性的过程有同样的结论。

当时间序列是高度持久（它们有单位根）的时候，我们在把它们直接用 370 于回归模型中时必须慎重（除非我们相信第10章的CLM假定成立）。使用水平形式的变量的一种替代方法是使用变量的一阶差分。对大多数高度持久的经济时间序列来说，一阶差分是弱相依的。使用一阶差分改变了模型的性质，但这种方法通常与水平形式的模型在提供的信息量上是相同的。当数据是高度持久的时候，我们通常更信赖一阶差分结果。在第18章，我们将谈到在多元回归分析中使用I（1）变量的几种最新的、更高级的方法。

当模型有完整动态性质，也就是方程中不再需要任何变量的其他滞后的时候，我们已经看到误差是序列不相关的。这一点非常有用，因为某些模型，比如自回归模型，被认为有完整动态性。在静态模型和有限分布滞后模型中，动态完整性假定经常不正确，这意味着误差是序列相关的。在第12章我们将看到如何来处理这类问题。

关键术语

渐近不相关
一阶自回归过程［AR(1)］
协方差平稳
动态完整模型
一阶差分
高度持久
一阶自积［I(1)］
零阶自积［I(0)］
一阶移动平均过程［MA(1)］

随机游走
有漂移的随机游走
序列不相关
平稳的AR(1)过程
平稳过程
强相依
趋势—平稳过程
单位根过程
弱相依非平稳过程

习 题

11.1 令$\{x_t: t=1, 2, \cdots\}$为协方差平稳过程，定义$\gamma_h = \mathrm{Cov}(x_t, x_{t+h})$，$h \geqslant 0$。［所以$\gamma_0 = \mathrm{Var}(x_t)$。］证明$\mathrm{Corr}(x_t, x_{t+h}) = \gamma_h / \gamma_0$。

11.2 令$\{e_t: t=-1, 0, 1, \cdots\}$为由独立同分布随机变量组成的序

列，它的均值为0，方差为1。定义以下的随机过程：

$$x_t = e_t - (1/2)e_{t-1} + (1/2)e_{t-2},\quad t = 1,2,\cdots$$

(i) 求出$\mathrm{E}(x_t)$和$\mathrm{Var}(x_t)$。它们中的哪个取决于t？

(ii) 证明$\mathrm{Corr}(x_t, x_{t+1}) = -1/2, \mathrm{Corr}(x_t, x_{t+2}) = 1/3$。[提示：最简单的方法是利用问题11.1中的公式。]

(iii) 在$h > 2$时，$\mathrm{Corr}(x_t, x_{t+h})$是多少？

(iv) $\{x_t\}$是渐近不相关过程吗？

11.3 假设时间序列过程$\{y_t\}$由$y_t = z + e_t$（$t = 1, 2, \cdots$）产生，$\{e_t\}$是均值为0，方差为σ_e^2的$i.i.d.$序列。随机变量z不随时间而变化，

371 它的均值为0，方差为σ_z^2。假定每个e_t都与z不相关。

11.4 令$\{y_t: t = 1, 2, \cdots\}$遵循式（11.20）那样的随机游走，且$y_0 = 0$。证明$\mathrm{Corr}(y_t, y_{t+h}) = \sqrt{t/(t+h)}$，其中$t \geqslant 1$，$h > 0$。

11.5 对于美国经济社会，令*gprice*代表一般价格水平的每月增长率，*gwage*代表每小时工资的每月增长率。[二者都是通过计算对数之差得到的：$gprice = \Delta\log(price), gwage = \Delta\log(wage)$。] 利用WAGEPRC.RAW中的月度数据，我们估计得到下面的分布滞后模型：

$$\begin{aligned} gprice = {} & -\underset{(0.000\,57)}{0.000\,93} + \underset{(0.052)}{0.119}\,gwage + \underset{(0.039)}{0.097}\,gwage_{-1} + \underset{(0.039)}{0.040}\,gwage_{-2} \\ & + \underset{(0.039)}{0.038}\,gwage_{-3} + \underset{(0.039)}{0.081}\,gwage_{-4} + \underset{(0.039)}{0.107}\,gwage_{-5} \\ & + \underset{(0.039)}{0.095}\,gwage_{-6} + \underset{(0.039)}{0.104}\,gwage_{-7} + \underset{(0.039)}{0.103}\,gwage_{-8} \\ & + \underset{(0.039)}{0.159}\,gwage_{-9} + \underset{(0.039)}{0.110}\,gwage_{-10} + \underset{(0.039)}{0.103}\,gwage_{-11} \\ & + \underset{(0.052)}{0.16}\,gwage_{-12} \\ & n = 273, R^2 = 0.317, \bar{R}^2 = 0.283 \end{aligned}$$

(i) 描述估计的滞后分布。*gwage*的哪一个滞后对*gprice*的影响最大？哪一个滞后的系数最小？

(ii) 哪些滞后的t统计量小于2？

(iii) 估计的长期倾向是多少？它与1的差距很大吗？解释在本例中LRP告诉了我们什么？

(iv) 你将用什么样的模型来直接求出LRP的标准误？

(v) 你将怎样检验再增加的6个*gwage*滞后的联合显著性？F分布的df是多少？（注意：你又失去了6组观测值。）

11.6 如果某人在时间$t-1$买入6个月国债，然后在时间t（3个月后）当做3个月国债卖出，令$hy6_t$代表他在这3个月持有期的收益率（以百分数计）。令$hy3_{t-1}$代表通过在时间$t-1$买入3个月国债而获得的持有收

益率。在时间 $t-1$，$hy3_{t-1}$是知道的。但是因为在时间 $t-1$ 还无法知道 $p3_t$（3 个月国债的价格），所以 $hy6_t$ 是不知道的。理性预期假说（expectation hypothesis，EH）认为，这两种不同的 3 个月投资的收益一般来讲应该是相同的。我们可以把这个结论用数学表达为一个条件期望：

$$\mathrm{E}(hy6_t \mid I_{t-1}) = hy3_{t-1}$$

372 式中，I_{t-1}在时间 $t-1$ 可观测得到的所有信息。因此，我们要估计模型：

$$hy6_t = \beta_0 + \beta_1 hy3_{t-1} + u_t$$

并检验 H_0：$\beta_1 = 1$。[我们也可以检验 H_0：$\beta_0 = 0$，但因为我们买入有不同到期日的资产常常会带来期间溢水（term premium），这会使得 $\beta_0 \neq 0$]。

（i）利用 INTQRT.RAW 中的数据（每 3 个月为一个时期），用 OLS 估计前面的方程，得到

$$\begin{aligned} \widehat{hy6}_t &= -0.058 + 1.104 hy3_{t-1} \\ &\quad\ (0.070) \quad (0.039) \\ &\quad\ n = 123, R^2 = 0.866 \end{aligned}$$

在 1% 的显著性水平上你拒绝 $\mathrm{H}_0:\beta_1 = 1$ 而取 $\mathrm{H}_0:\beta_1 \neq 1$ 吗？估计值看起来实际上异于 1 吗？

（ii）EH 的另一个含义是：一旦 $hy3_{t-1}$ 被确定下来，就没有其他时间为 $t-1$ 或更早的变量有助于解释 $hy6_t$。把 6 个月和 3 个月国债之间的差价的一个滞后添加到模型中，得到

$$\begin{aligned} \widehat{hy6}_t &= -0.123 + 1.053 hy3_{t-1} + 0.480(r6_{t-1} - r3_{t-1}) \\ &\quad\ (0.067) \quad (0.039) \qquad\quad (0.109) \\ &\quad\ n = 123, R^2 = 0.885 \end{aligned}$$

现在，$hy3_{t-1}$的系数是统计上异于 1 的吗？滞后的差价项是显著的吗？如果 $r6$ 在时间 $t-1$ 高于 $r3$，根据这个方程，你应该投资于 6 个月国债还是 3 个月国债？

（iii）$hy3_t$ 和 $hy3_{t-1}$之间的样本相关程度为 0.914。这为什么会使我们对前面的分析产生担心？

（iv）在（ii）估计的方程中，你怎样检验是否存在着季节性呢？

11.7 一个部分调整模型（partial adjustment model）如下：

$$\begin{aligned} y_t^* &= \gamma_0 + \gamma_1 x_t + e_t \\ y_t - y_{t-1} &= \lambda(y_t^* - y_{t-1}) + a_t \end{aligned}$$

式中，y_t^* 为 y 的理想或最优水平；y_t 为实际（观测到的）水平。举例来说，y_t^* 是某公司理想的库存增长率，x_t 是该公司销售的增长率。参数 γ_1 表示了 x_t 对 y_t^* 的影响程度。第二个方程描述了时间 t 的理想的 y 与时间 $t-1$ 的实际的 y 之间的关系决定了 y 是如何调整的。参数 λ（$0<\lambda<1$）度量了调整的速度。

(i) 将第一个方程中的 y_t^* 代入第二个方程，证明：我们可以得到

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t$$

373 并且，可以用 γ_j 和 λ 来表示 β_j，用 e_t 和 a_t 来表示 u_t。可见，由部分调整模型推出了一个有滞后的因变量和同期的 x 的模型。

(ii) 如果 $\mathrm{E}(e_t \mid x_t, y_{t-1}, x_{t-1}, \cdots) = \mathrm{E}(a_t \mid x_t, y_{t-1}, x_{t-1}, \cdots) = 0$，而且所有的序列都是弱相依的，你将怎样估计 β_j？

(iii) 如果 $\hat{\beta}_1 = 0.7$ 且 $\hat{\beta}_2 = 0.2$，那么 γ_1 和 λ 的估计值各是多少？

计算机习题

11.8 本题利用 HSEINV.RAW 中的数据。

(i) 求出 $\log(invpc)$ 的一阶自相关，然后再求 $\log(invpc)$ 在除掉线性趋势后的自相关。对 $\log(price)$ 作相同的计算。这两个序列中哪个可能有单位根？

(ii) 在 (i) 的基础上估计下面的方程：

$$\log(invpc_t) = \beta_0 + \beta_1 \Delta\log(price_t) + \beta_2 t + u_t$$

并按标准格式报告结果。对系数 $\hat{\beta}_1$ 作出解释，并判断它在统计上的显著性。

(iii) 除掉 $\log(invpc_t)$ 的线性趋势，然后在 (ii) 的回归方程中使用去掉了趋势的因变量（见 10.5 节），R^2 有什么变化吗？

(iv) 现在用 $\Delta\log(invpc_t)$ 作因变量。新的结果与 (ii) 中得到的有何不同？时间趋势还是显著的吗？为什么是或不是？

11.9 在例 11.7 中，用自然对数的变化来定义每小时工资和每小时产出的增长率：$ghrwage = \Delta\log(hrwage)$，$goutphr = \Delta\log(outphr)$。考虑式 (11.29) 的一个简单的扩展模型：

$$ghrwage_t = \beta_0 + \beta_1 goutphr_t + \beta_2 goutphr_{t-1} + u_t$$

它允许生产率增长的上升对工资的增长既有当期的影响又有滞后的影响。

(i) 利用 EARNS.RAW 中的数据估计这个方程，并用标准格式报告结果。$goutphr$ 的滞后值是统计上显著的吗？

(ii) 如果 $\beta_1 + \beta_2 = 1$，生产率增长的一个永久的上升将会在下一年里，完全反映到更高的工资增长水平上。相对于双侧对立假设检验 $\mathrm{H}_0: \beta_1 + \beta_2 = 1$。注意，最简单的方法是，像在第 10 章的例 10.4 中那样改写方程，使 $\theta = \beta_1 + \beta_2$ 直接出现在模型中。

(iii) 模型中需要 $goutphr_{t-2}$ 吗？说明理由。

11.10 (i) 在例 11.4 中，给定过去的收益时，时期 t 的期望收益有可能是 $return_{t-1}$ 的二次函数。为了检验这种可能性，利用 NYSE.RAW 中的数据估计

$$return_t = \beta_0 + \beta_1 return_{t-1} + \beta_2 return_{t-1}^2 + u_t$$

用标准格式报告结果。

374 (ii) 陈述并检验 $\mathrm{E}(return_t \mid return_{t-1})$ 不取决于 $return_{t-1}$ 这一虚拟假设。

[提示：这里要检验两个约束条件。]你有何结论？

(iii) 从方程中去掉 $return_{t-1}^2$，加进交互作用项 $return_{t-1}^2 \cdot return_{t-2}$，再来检验有效市场假设。

(iv) 你对在过去股票收益的基础上进行每周股票收益预测有何结论？

11.11 本题利用 PHILLIPS.RAW 中的数据。

(i) 在例 11.5 中，我们假定自然失业率是恒定的。在另一种形式的附加预期的菲利普斯曲线中，自然失业率受过去的失业水平影响。最简单的情况是，时期 t 的自然失业率与 $unem_{t-1}$ 相等。如果我们假定适应性预期，就会得到一个通货膨胀和失业率都是一阶差分形式的菲利普斯曲线：

$$\Delta inf = \beta_0 + \beta_1 \Delta unem + u$$

估计这个模型，用常见的格式报告结果，并讨论 $\hat{\beta}_1$ 的符号、大小和统计显著性。

(ii) 在式（11.9）和（i）中的模型这二者中，哪一个模型与数据更吻合？说明理由。

11.12 (i) 在方程（11.27）中添加一个线性时间趋势。在一阶差分方程中时间趋势是必要的吗？

(ii) 从式（11.27）中去掉时间趋势并添加变量 $ww2$ 和 $pill$（不要对虚拟变量进行差分）。这两个变量在 5% 的水平上是显著的吗？

(iii) 用（ii）中的模型估计 LRP 并求出它的标准误。把它们与从式（10.19）得出的结果进行比较，在式（10.19）中 gfr 和 pe 是以水平形式而不是差分形式出现的。

11.13 令 $inven_t$ 代表美国在 t 年的实际存货价值，GDP_t 代表实际国内生产总值，$r3_t$ 代表（过去的）3 个月国债的实际利率。根据过去情况得到的利率（大约）是 $r3_t = i3_t - inf_t$，这里的 $i3_t$ 是 3 个月国债的利率，inf_t 是年通货膨胀率 [见 Mankiw（1994，6.4 节）]。存货的变化 $\Delta inven_t$ 是当年的存货投资（inventory investment）。存货投资的加速模型（accelerator modle）为

$$\Delta inven_t = \beta_0 + \beta_1 \Delta GDP_t + u_t$$

式中，$\beta_1 > 0$。[参见 Mankiw（1994，17 章）等。]

(i) 利用 INVEN.RAW 中的数据估计这个加速模型。用常见的格式报告结果并解释方程的含义。$\hat{\beta}_1$ 是统计上大于零的吗？

(ii) 如果实际利率上升了，那么持有存货投资的机会成本上升，所以实际利率上升将导致存货的下降。把实际利率加进加速模型中并讨论得到的结果。实际利率的水平形式比它的一阶差分形式 $\Delta r3_t$ 更有效吗？

375 11.14 本题利用 CONSUMP.RAW 中的数据。一种消费的持久收入假说（permanent income hypothesis，PIH）认为，消费的增长是不可预测的。[还有一种 PIH 认为消费本身的变化是不可预测的；参见 Mankiw（1994，第 15 章）有关 PIH 的讨论。] 令 $gc_t = \log(c_t) - \log(c_{t-1})$ 表示人均实际（非耐用

品和服务的）消费的增长。那么 PIH 意味着$\mathrm{E}(gc_t \mid I_{t-1}) = \mathrm{E}(gc_t)$，这里的$I_{t-1}$表示在时间 $t-1$ 知道的信息，t 代表年份。

（i）估计$gc_t = \beta_0 + \beta_1 gc_{t-1} + u_t$ 以检验 PIH。明确表述虚拟假设和对立假设。你能得出什么结论？

（ii）在（i）的回归中添加变量 gy_{t-1}和 $i3_{t-1}$，其中 gy_t 是实际人均可支配收入的增长，$i3_t$ 是 3 个月国债的利率；注意，在回归中这二者都要加以滞后。添加的这两个变量是联合显著的吗？

11.15　本题利用 PHILLIPS.RAW 中的数据。

（i）估计失业率的AR(1)模型。用这个方程预测 1997 年的失业率。将它与 1997 年实际的失业率进行比较。（你可以从近年的总统经济报告中找到这个数据。）

（ii）在（i）的方程中加进通货膨胀的一期滞后。inf_{t-1}统计上显著吗？

（iii）利用（ii）中的方程预测 1997 年的失业率。这个结果比从（i）得出的结果更好些还是更糟些？

（iv）利用 6.4 节中的方法构造 1997 年失业率的一个 95% 的置信区间。1997 年的实际失业率在这个区间内吗？

第 12 章　时间序列回归中的序列相关和异方差

376　在本章，我们将讨论多元回归模型中误差项的序列相关这一重要问题。我们在第 11 章中了解到，如果一个模型的动态在适当的意义上被完整地设定了，它的误差就不会序列相关。所以，检验序列相关可以用来探测动态的错误设定。另外，静态模型和有限分布滞后模型经常有序列相关的误差，即使在模型没有被错误设定的时候也是这样。因此，很有必要了解在这些有用的模型中存在序列相关的后果和相应的补救方法。

在 12.1 节，我们讨论了在误差包含了序列相关时 OLS 的性质，在 12.2 节说明了如何检验序列相关。我们介绍了几种方法，其中一部分方法适用于有严格外生回归元的模型，另外一些方法对普通回归元（包括滞后因变量）都渐近有效。12.3 节阐述了在解释变量为严格外生的情况下，如何对序列相关进行补救的问题。12.4 节表明了使用差分过的数据是怎样常能消除误差的序列相关的。那么，如果出现了非常普通的序列相关，应该如何调整 OLS 标准误和统计量呢？12.5 节讨论了这方面的最新理论进展。

在第 8 章，我们讨论了横截面分析中异方差的检验和校正方面的问题。在 12.6 节，我们说明横截面情况下使用的方法怎样能扩展到时间序列分析中。这两种情况下的操作方法基本上是相同的，但有些由时间序列观测中时间上的相关性造成的细微区别，我们必须说清楚。另外，我们简单地提到了异方差的动态形式。

12.1 有序列相关误差的 OLS 性质

无偏性和一致性

在第 10 章我们证明了，在时间序列回归的前三个高斯-马尔科夫假定（TS.1～TS.3）的条件下，OLS 统计量是无偏的。特别是，定理 10.1 对误差是否序列相关没有作任何假定，因此，只要解释变量是严格外生的，无论误差的序列相关程度如何，实际上，$\hat{\beta}_j$ 都是无偏的。这与误差的异方差不会
377 造成 $\hat{\beta}_j$ 的偏误这一观察结果很相符。

在第 11 章，我们把严格外生性假定放松到 $\mathrm{E}(u_t \mid \boldsymbol{x}_t)=0$，并证明了当数据是弱相依的时候，$\hat{\beta}_j$ 的分布就紧缩成单一个点 β_j。实际上，$\hat{\beta}_j$ 仍然是一致的（但不一定无偏）。有关误差的序列相关的任何假定都不会改变这一结论。

效率和推断

因为高斯-马尔科夫定理（定理 10.4）要求误差既是同方差又是序列不相关的，那么，在存在序列相关的情况下 OLS 就不再是 BLUE 的了。更为重要的是，通常的 OLS 标准误和检验统计量不再生效，而且连渐近生效都称不上。这一点可以通过下面的例子来说明：让我们在前四个高斯-马尔科夫假定成立及误差遵循 AR（1）模型的情况下，来计算 OLS 估计量的方差。准确地说，我们假定

$$u_t = \rho u_{t-1} + e_t,\ t = 1,2,\cdots,n \tag{12.1}$$

$$|\rho| < 1 \tag{12.2}$$

式中，e_t 为均值为 0，方差为 σ_e^2 的不相关的随机变量。式（12.2）这个假定实际上就是第 11 章中的平稳性条件。

我们考虑简单的回归模型中的 OLS 斜率估计量的方差：

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

而且，为了简化这个公式，我们假定 x_t 的样本均值为 0（$\bar{x}=0$）。那么 β_1 的 OLS 估计量 $\hat{\beta}_1$ 可以表示为

$$\hat{\beta}_1 = \beta_1 + \mathrm{SST}_x^{-1} \sum_{t=1}^{n} x_t u_t \tag{12.3}$$

这里的$\mathrm{SST}_x=\sum_{t=1}^{n}x_t^2$。现在我们就来计算$\hat{\beta}_1$的方差（以$\mathbf{X}$为条件），但不再忽略序列相关的存在了：

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) &= \mathrm{SST}_x^{-2}\mathrm{Var}\Big(\sum_{t=1}^{n}x_tu_t\Big)\\ &= \mathrm{SST}_x^{2}\Big[\sum_{t=1}^{n}x_t^2\mathrm{Var}(u_t)+2\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}x_tx_{t+j}E(u_tu_{t+j})\Big] \quad (12.4)\\ &= \sigma^2/\mathrm{SST}_x+2(\sigma^2/\mathrm{SST}_x^2)\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\rho^jx_tx_{t+j}\end{aligned}$$

式中，$\sigma^2=\mathrm{Var}(u_t)$，我们利用了$\mathrm{E}(u_tu_{t+j})=\mathrm{Cov}(u_t,u_{t+j})=\rho^j\sigma^2$［见方程(11.4)］。方程中的第一项$\sigma^2/\mathrm{SST}_x$是当$\rho=0$时$\hat{\beta}_1$的方差，它也就是我们
378 熟悉的高斯-马尔科夫假定条件下的OLS方差。如果我们忽略序列相关，用平常的方法估计这个方差，当$\rho\neq0$时方差的估计量通常是有偏的，因为它忽略了式（12.4）中的第二项。在下面的例子中我们将看到，$\rho>0$是很常见的，在这种情况下，对所有的j都有$\rho^j>0$。此外，回归模型中的自变量在不同的时间经常是正相关的，以至于对于多数t和$t+j$来说，x_tx_{t+j}是正的。因此，在大多数的经济应用中，$\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\rho^jx_tx_{t+j}$是正的，所以通常的OLS方差公式$\sigma^2/\mathrm{SST}_x$低估了OLS估计量的真实方差。如果$\rho$很大或$x_t$有较高的正的序列相关——这是常见的情况——通常的OLS方差估计量的偏差会很大，我们会把OLS估计量想像得比实际上更为精确。

如果$\rho<0$，那么当j是奇数时ρ^j是负的，当j是偶数时ρ^j是正的，就很难断定$\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\rho^jx_tx_{t+j}$的符号。事实上，通常的OLS方差公式高估$\hat{\beta}_1$的真实方差也是有可能的。无论是在哪种情况下，通常的OLS估计量在有序列相关的时候对于$\mathrm{Var}(\hat{\beta}_1)$都是有偏的。

问题 12.1

假使u_t不遵循AR(1)模型而是遵循MA(1)模型$u_t=e_t+\alpha e_{t-1}$，试求出$\mathrm{Var}(\hat{\beta}_1)$，并证明如果$\alpha\neq0$，它就与通常的公式不同。

因为$\hat{\beta}_1$的标准误是$\hat{\beta}_1$的标准差的估计值，那么在有序列相关的时候使用通常的OLS标准误就不是可靠的了。所以，t统计量不能再用来检验单个假设。因为一个较小的标准误意味着一个较大的t统计量，所以$\rho>0$时通常的t统计量常常过大。用于检验多重假设的通常的F和LM统计量也是不可靠的。

出现滞后因变量时的序列相关

计量经济学的初学者经常受到警告：回归中出现滞后因变量时，误差有

序列相关的危险。几乎所有的计量经济学教材中都有以某种形式的以下表述："在模型中兼有滞后的因变量和序列相关的误差的时候，OLS是不一致的。"不幸的是，作为一般性的断言，这一表述是错误的。也存在着正确的说法，但一定要说得非常准确才行。

为了说明清楚，假使在给定 y_{t-1} 时 y_t 的期望值是线性的：

$$E(y_t|y_{t-1})=\beta_0+\beta_1 y_{t-1} \tag{12.5}$$

这里假定了平稳性，即 $|\beta_1|<1$。我们总可以把（12.5）写成

$$y_t=\beta_0+\beta_1 y_{t-1}+u_t \tag{12.6}$$

$$E(u_t|y_{t-1})=0 \tag{12.7}$$

这个模型满足OLS一致性所要求的关键假定TS.3′，因此OLS估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是一致的。很重要的是，如果没有进一步的假定，误差 $\{u_t\}$ 一定是序列 379 相关的。条件（12.7）保证了 u_t 与 y_{t-1} 不相关，但 u_t 和 y_{t-2} 却有可能相关。因为 $u_{t-1}=y_{t-1}-\beta_0-\beta_1 y_{t-2}$，那么 u_t 和 u_{t-1} 之间的协方差就是 $-\beta_1\mathrm{Cov}(u_t,y_{t-2})$，它并不一定为零。这样，虽然误差表现出序列相关性，模型也包含了一个滞后因变量，但OLS还是一致地估计了 β_0 和 β_1，因为它们是条件均值方程（12.5）中的参数。误差有序列相关性会导致OLS统计量不能可靠地用于检验，但它不会影响一致性。

那么，如果误差是序列相关的且回归元中包含一个滞后的因变量，在什么情况下OLS才是不一致的呢？如果完全按照式（12.6）那样，把模型写成误差形式，然后又和式（12.1）以及式（12.2）一样假定 $\{u_t\}$ 遵循一个平稳的AR(1)过程，即

$$E(e_t|u_{t-1},u_{t-2},\cdots)=E(e_t|y_{t-1},y_{t-2},\cdots)=0 \tag{12.8}$$

OLS就不是一致的。因为根据假定，e_t 与 y_{t-1} 不相关，所以 $\mathrm{Cov}(y_{t-1},u_t)=\rho\mathrm{Cov}(y_{t-1},u_{t-1})$ 不等于零，除非 $\rho=0$。由此可知，通过 y_t 对 y_{t-1} 的回归得到的 β_0 和 β_1 的OLS估计量不是一致的。

现在我们知道了，当误差 u_t 也遵循AR(1)模型时，对式（12.6）的OLS估计会导致不一致的估计量。但是，这样的表述虽然正确，仍不免误导人，我们不得不问：当误差符合AR(1)模型时，估计出的式（12.6）中的参数有什么意义呢？要想出所有的意义情况很难。但式（12.5）中那些参数至少告诉了我们在给定 y_{t-1} 时 y_t 的期望值。如果我们把式（12.6）和式（12.1）结合起来，就会发现，y_t 实际上遵循一个二阶自回归模型，即AR(2)模型。为了表明这点，我们写 $u_{t-1}=y_{t-1}-\beta_0-\beta_1 y_{t-2}$，把它代入 $u_t=\rho u_{t-1}+e_t$。于是，式（12.6）可写成

$$\begin{aligned}y_t&=\beta_0+\beta_1 y_{t-1}+\rho(y_{t-1}-\beta_0-\beta_1 y_{t-2})+e_t\\&=\beta_0(1-\rho)+(\beta_1+\rho)y_{t-1}-\rho\beta_1 y_{t-2}+e_t\\&=\alpha_0+\alpha_1 y_{t-1}+\alpha_2 y_{t-2}+e_t\end{aligned}$$

式中，$\alpha_0=\beta_0(1-\rho)$，$\alpha_1=\beta_1+\rho$，$\alpha_2=-\rho\beta_1$。又由于式（12.8），有

$$E(y_t \mid y_{t-1}, y_{t-2}, \cdots) = E(y_t \mid y_{t-1}, y_{t-2}) = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} \tag{12.9}$$

这个方程意味着给定 y 的所有过去值，y_t 的期望值取决于 y 的两期滞后。在包括预测在内的实际应用中，我们感兴趣的是方程(12.9)，在第 18 章将利用它进行预测。我们对参数 α_j 尤其感兴趣，因为在 AR(2)模型的适当的平稳性条件下——我们将在 12.3 节中介绍——对式(12.9)的 OLS 估计会产生 α_j 的一致的、渐近正态的估计值。

基本的问题是，我们需要为这个模型中既有一个滞后因变量又有序列相关的误差找到一个合适的理由。动态模型的误差的序列相关，通常是由动态回归函数没有被完整设定造成的：在前一个例子中，我们应该在方程中添加 y_{t-2}。

380 在第 18 章，我们将看到几个例子，在它们的模型中，有滞后的因变量，误差是序列相关的，而且误差也与 y_{t-1} 相关。但是，即使在这些情况下，误差也不是一个自回归过程。

12.2 序列相关的检验

在本节，我们介绍在下面的多元线性回归模型中检验误差项的序列相关性的几种方法：

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t$$

我们首先考虑回归元是严格外生的情况。回忆一下，它需要误差 u_t 与所有时期的回归元都不相关（见 10.3 节），因此，它排除了含有滞后的因变量的那些模型。

回归元为严格外生时对AR(1)序列相关的 t 检验

多元回归模型中的误差项的序列相关可以有很多种形式，最常见的形式——也是最容易解决的——是由方程（12.1）和方程（12.2）给出的AR(1)模型。在前面一节，我们解释了，当误差一般地序列相关时，运用 OLS 将意味着什么，也推导了一个简单的 AR（1）形式的回归模型中的 OLS 斜率估计量的方差。我们现在介绍如何检验AR(1)序列相关的存在。所作虚拟假设是不存在序列相关；于是，就像在检验异方差时一样，我们假定最理想的情况，并要求数据提供有力的证据，来表明无序列相关这一理想的假定不成立。

我们先推导一个大样本检验方法，它是以假定解释变量是严格外生的为条件的：给定自变量的过去的所有信息，u_t 的期望值为零。除此之外，在

式（12.1）中我们必须假定：

$$\mathrm{E}(e_t \mid u_{t-1}, u_{t-2}, \cdots) = 0 \tag{12.10}$$

和

$$\mathrm{Var}(e_t \mid u_{t-1}) = \mathrm{Var}(e_t) = \sigma_e^2 \tag{12.11}$$

这些是AR(1)模型的标准假定（当 $\{e_t\}$ 是 $i.i.d.$ 序列时成立），它们使得我们可以把第11章的大样本结果应用于动态回归中。

如同异方差的检验那样，虚拟假设就是相应的高斯-马尔科夫假定为真。在AR（1）模型中，这个虚拟解释应该是，误差是序列不相关的，即

$$\mathrm{H}_0: \rho = 0 \tag{12.12}$$

怎样来检验这个假设呢？如果 u_t 被观测到了，那么，在式(10.10)和式(10.11)条件下，我们可以把定理11.2的渐近正态结果直接应用于动态回归模型：

$$u_t = \rho u_{t-1} + e_t, \quad t = 2, \cdots, n \tag{12.13}$$

381 [在虚拟假设 $\rho = 0$ 下，$\{u_t\}$ 显然是弱相依的。]换句话说，对所有的 $t = 2, \cdots, n$，我们可以通过做 u_t 对 u_{t-1} 无截距的回归来估计 ρ，而且还可以使用 $\hat{\rho}$ 的通常的 t 统计量。但是，这个做法是行不通的，因为 u_t 无法观测到。可是，就像对异方差的估计那样，我们可以用相应的OLS残差 $\hat{u}_t$ 来代替 u_t。因为 $\hat{u}_t$ 取决于估计量 $\hat{\beta}_0$，$\hat{\beta}_1$，…，$\hat{\beta}_k$，在回归中用 $\hat{u}_t$ 代替 u_t 对 t 统计量的分布并非没有影响。幸运的是，由于严格外生性假定，t 统计量的大样本分布不受OLS残差代替误差的影响。这个结论的证明超出了本书的范围，可参见Wooldridge（1991b）。

我们可以简单地总结一下AR(1)序列相关的渐近检验。

回归元严格外生时AR(1)序列相关的检验

(i) 做 y_t 对 x_{t1}，…，x_{tk} 的OLS回归，得到OLS残差 $\hat{u}_t$（$t = 1, 2, \cdots, n$）。

(ii) 运行回归：

$$\hat{u}_t \text{ 对 } \hat{u}_{t-1}, \quad t = 2, \cdots, n \tag{12.14}$$

得到 $\hat{u}_{t-1}$ 的系数 $\hat{\rho}$ 和它的 t 统计量 $t_{\hat{\rho}}$。（这个回归可以包含也可以不包含截距；$\hat{\rho}$ 的 t 统计量会受到轻微的影响，但在两种情况下的 t 统计量都是渐近生效的。）

(iii) 按照通常的方法，用 $t_{\hat{\rho}}$ 去检验 H_0：$\rho = 0$ 和 H_1：$\rho \neq 0$。（实际上，因为 $\rho > 0$ 往往是更容易出现的，故也可采用对立假设 H_1：$\rho > 0$。）一般地，只有 H_0 在5%的水平上被拒绝时，我们才认为存在着需要处理的序列相关。和以前一样，我们最好报告检验得到的 p 值。

在判断是否有必要考虑序列相关的问题时，我们应该注意实际显著性和

统计显著性的不同。如果样本容量较大，即使 ρ 实际上很小，我们还是有可能发现序列相关的；当 ρ 接近于零时，通常的 OLS 推断程序基本上也都不至于太离谱［见方程（12.4）］。当然，由于时间序列数据集通常较小，这样的情况在时间序列应用中较为少见。

例 12.1　菲利普斯曲线中 AR（1）序列相关的检验

在第 10 章我们估计了一个静态的菲利普斯曲线，它解释了美国的通货膨胀和失业之间的替代关系（见例 10.1）。在第 11 章，我们研究了一种特殊的附加预期的菲利普斯曲线，并作了适应性预期的假定（见例 11.5）。现在，我们来检验这两个方程中是否有序列相关。因为附加预期的曲线用了 $\Delta inf_t = inf_t - inf_{t-1}$ 作因变量，就又少了一次观测。

382 在静态菲利普斯曲线中，式（12.14）中的回归结果为：$\rho = 0.573$，$t = 4.93$，而 p 值 $= 0.000$（48 次观测）。这些结果比较充分地证明了正的、一阶序列相关的存在。它的一个后果是，第 10 章的标准误和 t 统计量不再适用。与此形成对比的是，在附加预期的菲利普斯曲线中检验AR(1)序列相关的结果是：$\rho = -0.036$，$t = -0.297$，而 p 值 $= 0.775$（47 次观测）；也就是说，没有证据表明在附加预期的菲利普斯曲线中有AR(1)序列相关。

尽管式（12.14）中的检验方法是从 AR（1）模型推导出来的，但这种检验方法也可以用于探测其他类型的序列相关。记住，ρ 是 u_t 和 u_{t-1} 之间的相关性的一致估计量。能够造成相邻误差项相关的任何序列相关都可以用这种检验方法探测出来。但是，它无法探测出相邻的误差不相关，即 $\text{Corr}(u_t, u_{t-1}) = 0$ 时的序列相关。（例如 u_t 和 u_{t-2} 可能是相关的。）

问题 12.2　你将怎样用回归式（12.14）来构造 ρ 的近似 95% 的置信区间？

在使用式（12.14）中的 t 统计量的过程中，我们必须假定式（12.13）中的误差满足相应的同方差假定（12.11）。实际上，很容易做到让这种检验对 e_t 的异方差情形是稳健的：只要用第 8 章的通常的异方差—稳健 t 统计量就行了。例 12.1 中的静态菲利普斯曲线异方差-稳健 t 统计量是 4.03，它比非稳健性 t 统计量要小，但仍然非常显著。在 12.6 节，我们将进一步讨论时间序列回归中的异方差问题，其中包括它的动态形式。

经典假定条件下的德宾-沃森统计量

AR(1)序列相关的另一种检验方法是德宾-沃森检验（Durbin-Watson test）。**德宾-沃森统计量**［Durbin-Watson（DW）statistic］也是以 OLS 残差为

基础的：

$$DW=\frac{\sum_{t=2}^{n}(\hat{u}_t-\hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2} \tag{12.15}$$

简单的代数运算表明 DW 和式（12.14）中的 $\hat{\rho}$ 有近似的关系：

$$DW\approx 2(1-\hat{\rho}) \tag{12.16}$$

二者的上述关系只是近似的关系而非准确的关系，一个原因是，$\hat{\rho}$ 的分母中只有 $\sum_{t=2}^{n}\hat{u}_{t-1}^2$，而 DW 统计量的分母是所有 OLS 残差的平方和。即使样本的规模不是很大，式（12.6）的等式两边也很接近。因此，基于 DW 的检验和基于 $\hat{\rho}$ 的 t 检验在概念上是一样的。

383 德宾和沃森（Durbin and Watson，1950）推导出了 DW 的（以 $\mathbf{X}$ 为条件的）分布，他们的推导是以全套经典线性模型假定为条件的，包括误差项的正态性。不幸的是，这个分布取决于自变量的值。（它还取决于样本规模、回归元的数量和回归是否包含截距等。）一些计量经济学软件包列出了 DW 的临界值和 p 值，但很多软件包并不提供这些数据。无论在什么情况下，这些数据都依赖于全套 CLM 假定。

有若干计量经济学课本报告临界值的上下界，这些数值依赖于所取的置信区间、对立假设、观测的次数和回归元的个数。（我们假定模型中包含截距。）通常情况下，DW 检验的计算用于检验对立假设：

$$\mathrm{H}_1: \rho>0 \tag{12.17}$$

由于式(12.16)中的近似关系，$\hat{\rho}\approx 0$ 意味着 $DW\approx 2$，而 $\hat{\rho}>0$ 意味着 $DW<2$。因此，为了拒绝虚拟假设（12.12）而接受假设（12.17），我们希望找到显著地小于 2 的一个 DW 的值。不幸的是，由于在获取 DW 在虚拟假设下的分布方面有困难，我们必须将 DW 与两组临界值进行比较。这些临界值通常被表示为 d_U（上界）和 d_L（下界）。如果 $DW<d_L$，拒绝 H_0 而取假设（12.17）；如果 $DW>d_U$，不能拒绝 H_0。如果 DW 在二者之间，则无明确的结论。

举个例子，如果我们选择显著性水平为 5%，$n=45$，$k=4$，$d_U=1.720$，$d_L=1.336$ [见 Savin 和 White（1997)]。如果 $DW<1.336$，我们在 5% 的水平上拒绝没有序列相关的虚拟假设；如果 $DW>1.72$，我们不能拒绝 H_0；如果 $1.336\leqslant DW\leqslant 1.72$，我们无法从检验中得出结论。

在静态菲利普斯曲线的例子 12.1 中，经计算 $DW=0.80$。我们可以求出 $k=1$，$n=50$ 时 1% 的下界值（Savin and White，1977）：$d_L=1.32$。因此，我们在 1% 的水平上拒绝序列相关为零的假设，认为存在正的序列相关。（利用以前的 t 检验，我们可以断定 p 值的前 3 个小数位上都是零。）对于附加预期的菲利普斯曲线，$DW=1.77$，它正好落在 5% 水平上的接受区域（$d_U=1.59$）。

一个精确的 DW 抽样分布可以列表表示出来，这是 DW 对于式（12.14）中的 t 检验的惟一优点。但是，所列出的临界值只有在所有的 CLM 假定都得到满足的情况下才是有效的，而且它们可能会导致很宽的不确定区域。考虑到这一点，DW 实际上的劣势就很大了。而式（12.14）中得到的 t 统计量计算起来简单，而且即使误差不是正态分布的，它也是渐近有效的。如果存在着取决于 x_{tj} 的异方差，t 统计量也是有效的，并且很容易使它适用于任何形式的异方差。

回归元不是严格外生时AR(1)序列相关的检验

当解释变量不是严格外生的时候，会有一个或更多的 x_{tj} 与 u_{t-1} 相关，回归式（12.14）中的 t 统计量和德宾-沃森统计量都无效，即使在大样本的情况下也是如此。非严格外生回归元的一个重要的例子是模型中包含滞后因变量的情况：y_{t-1} 和 u_{t-1} 显然是相关的。德宾（Durbin，1970）提出了两种
384 代替 DW 统计量的办法，这两种办法适用于模型中包含一个滞后的因变量且其他回归元都是非随机（或者，更一般地，严格外生）的情况。第一种被称为德宾 h 统计量（Durbin's h statistic）。这个统计量有个缺点，就是它不是总能计算出来，所以这里不打算介绍。

德宾的另一个统计量计算起来比较简单，而且不论有多少个非严格外生解释变量，它都是有效的。如果解释变量恰好是严格外生的，这个检验方法同样有效。

一般回归元的序列相关的检验

(i)做 y_t 对 $x_{t1},\cdots,x_{tk}$ 的回归，求出 OLS 残差 $\hat{u}_t(t=1,2,\cdots,n)$。

(ii)做回归：

$$\hat{u}_t \text{ 对 } x_{t1},x_{t2},\cdots,x_{tk},\ \hat{u}_{t-1},\quad t=2,\cdots,n \tag{12.18}$$

求得 $\hat{u}_{t-1}$ 的系数 $\hat{\rho}$ 及它的 t 统计量 $t_{\hat{\rho}}$。

(iii) 按照平常的方法，用 $t_{\hat{\rho}}$ 去检验 $\mathrm{H}_0:\rho=0$ 和 $\mathrm{H}_1:\rho\neq 0$（或者用单侧对立假设）。

在方程（12.18）中，我们做 OLS 残差对所有自变量的回归，其中包括截距和滞后的残差。滞后残差的 t 统计量是检验在 AR（1）模型（12.13）［当我们在 H_0 下增加条件 $\mathrm{Var}(u_t|\boldsymbol{x}_t,u_{t-1})=\sigma^2$ 时］中式（12.12）是否成立的一种有效的方法。x_{tj} 中可以含有任何个数的滞后因变量，同时，也允许存在非严格外生性解释变量。

把 x_{t1}，…，x_{tk} 加进回归方程中，明显考虑到了每个 x_{tj} 与 u_{t-1} 之间的相关性，这一做法也使得 $t_{\hat{\rho}}$ 在大样本情况下有渐近的 t 分布。式（12.14）中的 t 统计量忽略了 x_{tj} 和 u_{t-1} 之间可能的相关性，所以在回归元不是严格外生的情况下它不是有效的。顺便提一下，由于 $\hat{u}_t=y_t-\hat{\beta}_0-\hat{\beta}_1x_{t1}-\cdots-\hat{\beta}_kx_{tk}$，

可以证明如果把 y_t 放在 $\hat{u}_t$ 的位置上作为式（12.18）中的因变量，$\hat{u}_{t-1}$ 的 t 统计量不变。

容易使式（12.18）中的 t 统计量对未知形式的异方差具有稳健性［特别是当 $\mathrm{Var}(u_t|x_t,u_{t-1})$ 不是常数时］：只要用 $\hat{u}_{t-1}$ 的异方差-稳健 t 统计量就行了。

例 12.2　检验最低工资方程中的 AR（1）序列相关

在第 10 章（见例 10.9），考察了最低工资对波多黎各就业率的影响。我们现在来检验误差中是否包含了序列相关，检验过程中并不假定最低工资和 GNP 有严格外生性。［就像在习题 10.9 中那样，我们在方程（10.38）中
385 加入波多黎各实际 GNP 的对数值］。在这里，我们假定潜在的随机过程是弱相依的，但允许它们包含线性时间趋势（在回归中引进 t）。

令 $\hat{u}_t$ 代表 OLS 残差，使用现有的 37 次观测，运行下面的回归：

$$\hat{u}_t \text{ 对 } \log(mincov_t),\log(prgnp_t),\log(usgnp_t),t \text{ 和 } \hat{u}_{t-1}$$

估计得到的 $\hat{u}_{t-1}$ 的系数是 $\hat{\rho}=0.481$，$t=2.89$（双侧 p 值 $=0.007$）。因此，有足够证据表明误差中有 AR(1) 序列相关，这意味着我们得到的 $\hat{\beta}_j$ 的 t 统计量不能用于推断。但是别忘了，如果 u_t 与每个解释变量是同期不相关的，$\hat{\beta}_j$ 仍然是一致的。碰巧，如果我们使用回归式（12.14），就会得到 $\hat{\rho}=0.417$，$t=2.63$，在这种情况下检验的结果大体相似。

更高阶序列相关的检验

式（12.18）中的检验方法很容易就可以扩展到更高阶序列相关的情况。举个例子，假使我们想检验

$$\mathrm{H}_0: \rho_1=0,\ \rho_2=0 \tag{12.19}$$

在下面的 AR(2) 模型

$$u_t=\rho_1 u_{t-1}+\rho_2 u_{t-2}+e_t$$

中，这个另选的序列相关模型使得我们可以检验二阶序列相关。和以前一样，我们用 OLS 方法求出 OLS 残差 $\hat{u}_t$。接着，我们运行回归：

$$\hat{u}_t \text{ 对 } x_{t1},x_{t2},\cdots,x_{tk},\hat{u}_{t-1} \text{ 和 } \hat{u}_{t-2},\quad t=3,\cdots,n$$

以得到 $\hat{u}_{t-1}$ 和 $\hat{u}_{t-2}$ 的联合分布的 F 统计量。如果这两个滞后在足够小的水平上，比如 5%，是联合显著的话，我们就拒绝式（12.19），认为误差是序列相关的。

更一般地看，我们可以在 q 阶自回归模型中检验序列相关

$$u_t=\rho_1 u_{t-1}+\rho_2 u_{t-2}+\cdots+\rho_q u_{t-q}+e_t \tag{12.20}$$

对应的虚拟假设是

$$H_0: \rho_1 = 0,\ \rho_2 = 0, \cdots, \rho_q = 0 \qquad (12.21)$$

AR（q）序列相关的检验

（i）做 y_t 对 x_{t1}、…，x_{tk}（$t=1$、2，…、n）的 OLS 回归，求出 OLS 残差。

（ii）做回归：

386 $$\hat{u}_t \text{ 对 } x_{t1}, x_{t2}, \cdots, x_{tk}, \hat{u}_{t-1}, \hat{u}_{t-2}, \cdots, \hat{u}_{t-q}, t=(q+1), \cdots, n \qquad (12.22)$$

（iii）计算式（12.22）中的 $\hat{u}_{t-1}, \hat{u}_{t-2}, \cdots, \hat{u}_{t-q}$ 联合显著性的 F 检验。[也可以把 y_t 作为式（12.22）中的因变量来计算 F 统计量，结果是相同的。]

如果假定 x_{tj} 随机严格外生，以使 x_{tj} 与 u_{t-1}，u_{t-2}，…，u_{t-q} 不相关，那么，x_{tj} 可以从式（12.22）中省略掉。在回归中包含 x_{tj}，可以使得无论有没有严格外生性假定检验都是有效的。这个检验需要同方差假定：

$$\mathrm{Var}(u_t \mid \boldsymbol{x}_t, u_{t-1}, \cdots, u_{t-q}) = \sigma^2 \qquad (12.23)$$

我们可以参照第 8 章描述的方法计算严格异方差—稳健性 F 统计量。

一种可以代替 F 检验的方法是使用拉格朗日乘数（Lagrange multiplier，*LM*）形式的统计量。（我们在第 5 章的横截面分析中介绍过用于检验排除性约束的 *LM* 统计量。）检验式（12.21）的 *LM* 统计量是

$$LM = (n-q) R_{\hat{u}}^2 \qquad (12.24)$$

式中，$R_{\hat{u}}^2$ 是回归式（12.22）中的常见的 R-平方。在虚拟假设下，$LM \overset{a}{\sim} a\chi_q^2$。它通常被称做AR($q$)序列相关的**布劳殊-戈弗雷检验**（Breusch-Godfrey test）。*LM* 统计量还要求式（12.23），但是，可以经过变换使它变得对异方差是稳健的。[详尽的分析参见 Wooldridge（1991b）。]

例 12.3 AR（3）序列相关的检验

在氯化钡工业的事件研究中（见例 10.5），使用了月度数据，因此我们可以用它们来检验是否存在更高阶的序列相关。我们检验方程（10.22）中的误差的 AR（3）序列相关，以此说明方法。利用回归式（12.22），$\hat{u}_{t-1}$，$\hat{u}_{t-2}$ 和 $\hat{u}_{t-3}$ 的联合分布的 F 统计量是 $F=5.12$。在最开始的时候，我们有 $n=131$，而在扩充的回归式（12.22）中失去了三次观测。因为在本例中估计 10 个参数，F 统计量的 df 是 3 和 118。F 统计量的 p 值是 0.002 3，所以有足够的证据表明AR(3)序列相关的存在。

对于没有经过季节性调整的季度或月度数据，我们有时希望能够检验季节形式的序列相关。例如，对于季度数据，我们可以假定下面的自回归模型

$$u_t = \rho_4 u_{t-4} + e_t \tag{12.25}$$

通过前面的AR（1）序列相关的检验，我们对检验的程序应该很清楚了。当回归元是严格外生时，我们可以在下面的回归中使用 $\hat{u}_{t-4}$的 t 检验：

387

$$\hat{u}_t \text{ 对 } \hat{u}_{t-4}, \quad t = 5, \cdots, n$$

也可以使用修正的德宾-沃森统计量［见 Wallis（1972）］。当 x_{tj}不是严格外生的时候，我们可以用式（12.18）中的回归，只要用 $\hat{u}_{t-4}$替代 $\hat{u}_{t-1}$就行了。

在例12.3中，数据是月度的，而且未经过季节性调整，因此，检验 u_t 和u_{t-12}之间的相关性就是有必要的。做 $\hat{u}_t$ 对 $\hat{u}_{t-12}$的回归，得到 $\hat{\rho}_{12} = -0.187$，p 值$=0.028$，这表明了负的季节自相关的存在。（把回归元也包括进来只稍微改变了结果：$\hat{\rho}_{12} = -0.170$，p 值$=0.052$。）这有点不正常，但我们找不到比较明确的原因。

问题12.3☞ 在有季度数据的情况下，假使你想检验是否有一阶或四阶序列相关。如果回归元是外生的，你将如何操作？

12.3 对严格外生回归元的序列相关的校正

如果用12.2节中的检验方法探测到了序列相关的存在，我们就必须对此采取一定的措施。如果我们的目标是估计一个有完整动态的模型，就要重新设定模型。在实际应用中，我们的目标可能并不是要估计一个完整动态模型，而是要找到一个执行统计推断的办法：正如在12.1节看到的，常见的OLS检验统计量不再有效。在这一节，我们从AR（1）序列相关这种重要的情形开始来探讨这方面的问题。解决这个问题的传统办法是假定回归元是固定的，而实际上我们真正需要的是假定回归元是严格外生的。所以，至少在解释变量包含滞后因变量时，我们不应该使用这些校正方法。

在AR(1)模型中求最优线性无偏估计量

我们假定高斯-马尔科夫假定TS.1～TS.4都成立，但放宽假定TS.5。而且，假设误差遵循AR(1)模型：

$$u_t = \rho u_{t-1} + e_t, \quad t = 1, 2, \cdots \tag{12.26}$$

别忘了，假定TS.2要求以 $\boldsymbol{X}$ 为条件的 u_t 均值为零。在下面的分析中，我们把 $\boldsymbol{X}$ 这一条件写出来，以简化表述。于是，将 u_t 的方差写为

$$\mathrm{Var}(u_t)=\frac{\sigma_e^2}{1-\rho^2} \qquad (12.27)$$

为简洁起见．考虑只有一个解释变量的情况：

$$y_t=\beta_0+\beta_1 x_t+u_t,\quad t=1,2,\cdots,n$$

既然这个方程的麻烦在于 u_t 中存在的序列相关，那么，对方程进行变换，从而达到消除序列相关的目的就很有必要。当 $t\geqslant 2$ 时，有

388

$$\begin{aligned}y_{t-1}&=\beta_0+\beta_1 x_{t-1}+u_{t-1}\\ y_t&=\beta_0+\beta_1 x_t+u_t\end{aligned}$$

现在，我们把第一个方程两边都乘以 ρ，然后再从第二个方程把它减去，得到

$$y_t-\rho y_{t-1}=(1-\rho)\beta_0+\beta_1(x_t-\rho x_{t-1})+e_t,\ t\geqslant 2$$

式中，设 $e_t=u_t-\rho u_{t-1}$。上式可以写为

$$\tilde{y}_t=(1-\rho)\beta_0+\beta_1\tilde{x}_t+e_t,\quad t\geqslant 2 \qquad (12.28)$$

式中

$$\tilde{y}_t=y_t-\rho y_{t-1},\quad \tilde{x}_t=x_t-\rho x_{t-1} \qquad (12.29)$$

被称为**准差分数据**(quasi-differenced data)。(若 $\rho=1$，它们是差分数据，但我们假定了 $|\rho|<1$。)式(12.28)中的误差项是序列不相关的；实际上，这个方程满足所有的高斯–马尔科夫假定。这意味着，如果我们知道了 ρ，就可以通过做 $\tilde{y}_t$ 对 $\tilde{x}_t$ 的回归估计出 β_0 和 β_1，当然要把估计到的截距除以 $(1-\rho)$。

来自于式（12.28）的 OLS 统计量还不完全是 *BLUE*，因为我们没有利用第一个时期的数据。这个问题很容易解决，把 $t=1$ 时的方程表示为

$$y_1=\beta_0+\beta_1 x_1+u_1 \qquad (12.30)$$

既然每个 e_t 与 u_1 都不相关，我们可以把式（12.30）加进式（12.28）中，仍然保持了误差的序列不相关。但是，根据式（12.27），$\mathrm{Var}(u_1)=\sigma_e^2/(1-\rho^2)\sigma_e^2=\mathrm{Var}(e_t)$。［当 $|\rho|\geqslant 1$ 时，方程（12.27）显然不成立，这也是我们假定平稳性条件的原因所在。］这样，我们必须把方程（12.30）两边都乘以 $(1-\rho^2)^{1/2}$，以使误差有相同的方差：

$$(1-\rho^2)^{1/2}y_1=(1-\rho^2)^{1/2}\beta_0+\beta_1(1-\rho^2)^{1/2}x_1+(1-\rho^2)^{1/2}u_1$$

或

$$\tilde{y}_t=(1-\rho^2)^{1/2}\beta_0+\beta_1\tilde{x}_1+\tilde{u}_1 \qquad (12.31)$$

这里，$\tilde{u}_1=(1-\rho^2)^{1/2}u_1,\tilde{y}_1=(1-\rho^2)^{1/2}y_1$，等等。式（12.31）中的误差的方差为 $\mathrm{Var}(\tilde{u}_1)=(1-\rho^2)\mathrm{Var}(u_1)=\sigma_e^2$，于是可以把式（12.28）和式（12.31）一并用在 OLS 回归中。以上就给出了在假定 TS.1～TS.4 成立 u_t 遵循AR(1)模型的条件下，β_0 和 β_1 的 BLUE 估计量。这又是一个广义最小二乘法（GLS）估计量的例子。我们在第 8 章看到过异方差条件下的 *GLS*

估计量。

在回归中加进更多的回归元，情况大致相同。当 $t\geq 2$ 时，我们利用方程

$$\tilde{y}_t = (1-\rho)\beta_0 + \beta_1\tilde{x}_{t1} + \cdots + \beta_k\tilde{x}_{tk} + e_t \tag{12.32}$$

389 式中，$\tilde{x}_{tj} = x_{tj} - \rho x_{t-1,j}$。当 $t=1$ 时，我们有 $\tilde{y}_1 = (1-\rho^2)^{1/2}y_1$，$\tilde{x}_{1j} = (1-\rho^2)^{1/2}x_{1j}$，而截距为 $(1-\rho^2)^{1/2}\beta_0$。对于给定的 ρ，对数据进行变换并施行 OLS，做起来很简单。除非 $\rho=0$，否则 GLS 统计量，也就是对变型后的数据施行 OLS 得到的估计量，一般与最初未经变换得到的 OLS 统计量是不同的。GLS 统计量是 BLUE 的，而且因为变换后方程中的误差是序列不相关和同方差的，从变换后的方程中得出的 t 统计量和 F 统计量都是正确的（至少是渐近正确的，如果误差 e_t 是正态分布的，就确确实实地正确了）。

有AR（1）误差的可行GLS估计

GLS 估计量存在的问题是，实际上我们很少能够知道 ρ。不过，我们已经知道了如何取得 ρ 的一致估计值：与在方程（12.14）中完全一样，只要做 OLS 残差对它们相应的滞后值的回归就行了。接下来，我们用估计值 $\hat{\rho}$ 来代替 ρ，来获得准差分变量。然后，对下面的方程应用 OLS：

$$\tilde{y} = \beta_0\tilde{x}_{t0} + \beta_1\tilde{x}_{t1} + \cdots + \beta_k\tilde{x}_{tk} + 误差项 \tag{12.33}$$

式中，$\tilde{x}_{t0} = (1-\hat{\rho})$，$t\geq 2$，而 $\tilde{x}_{10} = (1-\hat{\rho}^2)^{1/2}$。这将给出 β_j 的**可行广义最小二乘法**（feasible GLS，FGLS）估计量。式（12.33）中的误差项包含 e_t，也包含涉及 $\hat{\rho}$ 的估计误差的那些项。幸运的是，$\hat{\rho}$ 的估计误差并不影响 FGLS 估计量的渐近分布。

AR（1）模型的可行的 GLS 估计

（i）做 y_t 对 x_{t1}，…，x_{tk} 的 OLS 回归，求出 OLS 残差 $\hat{u}_t$（$t=1, 2, \cdots, n$）。

（ii）做方程（12.14）的回归，求出 $\hat{\rho}$。

（iii）用 OLS 估计方程（12.33）中的 β_0，β_1，…，β_k。常见的标准误、t 统计量和 F 统计量都是渐近正确的。用 $\hat{\rho}$ 来代替 ρ 的代价是，可行的 GLS 估计量失去了易于处理的有限样本性质。特别是当数据是弱相依的时候，它就不再是无偏的了，尽管仍然是一致的。而且，即使式（12.32）中的 e_t 是正态分布的，t 统计量和 F 统计量也只是渐近 t 分布和渐近 F 分布的，这是因为存在着 $\hat{\rho}$ 的估计误差。对于大多数目的而言，这样的结果已经很不错了，尽管在样本容量较小时我们要慎重一些。

既然 FGLS 估计量不是无偏的，我们当然不能说它 BLUE。尽管如此，当序列相关的 AR（1）模型成立（而且解释变量为严格外生）时，它还是比 OLS 估计量更渐近有效。这个结论也是假定了这个时间序列是弱相依的。

AR（1）模型的FGLS估计有很多种名称，根据估计 ρ 的方法的不同和处理第一次观测的方法的不同而不同。**科克伦-奥克特估计**［Cochrane-Orcutt（CO）estimation］省略了第一次观测，用的是式（12.14）中的 $\hat{\rho}$，而**普莱斯-温斯登估计**［Prais-Winsten（PW）estimation］则采用前面讲的方法使用了第一次观测。大致说来，是否利用第一次观测并不会带来很大差别，但因为很多时间序列的样本很小，实际应用中的这种差别可能较大。

390 在实际计算过程中，一个迭代过程可能既要用到科克伦-奥克特方法又要用到普莱斯-温斯登方法。一旦用式（12.14）中的 $\hat{\rho}$ 求出FGLS估计值，我们就可以计算出一组新的残差，从式（12.14）中求出 ρ 的新的估计值，并用新的估计值 $\hat{\rho}$ 对数据进行变换，再用OLS来估计式（12.33）。我们可以把这个过程重复很多次，直到 ρ 的估计值与上一次的估计值差别很小为止。很多回归程序包能够自动执行不断迭代的程序，所以不会给我们带来任何额外的负担。但是，很难说多进行几次迭代计算就更好一些，在有些情况下，可能会好些，但从理论上来说，反复计算得到的估计值与只用第一次计算得到的估计值，它们的大样本性质是一样的。若想了解关于这些及其他方法的详细内容，请参见Davidson和MacKinnon（1993，第10章）。

例12.4　事件研究中的科克伦-奥克特估计

我们用迭代的科克伦-奥克特估计来估计例10.5中的方程。为了便于比较，我们也把OLS的结果列在表12.1中。

科克伦-奥克特估计中统计上显著的系数与OLS估计值没有太大差别［特别是log(*chempi*)和log(*rtwex*)的系数和 *afdec*6］。统计上不显著的系数，由于采用方法的不同而有所不同，并不令人奇怪，并且这种差别可能很大。

第一栏里的标准误比第二栏里相应的标准误都要高些，不知道你注意到没有，其实这很正常，因为科克伦-奥克特标准误考虑到了序列相关，而OLS标准误没有考虑。正如我们在12.1节谈到的，OLS标准误通常会低估OLS估计量的真实抽样方差，而且在有严重的序列相关的时候，我们无法信赖OLS标准误。所以，国际贸易委员会的决定对中国进口的影响，现在看来没有以前我们所认为的那么显著（$t_{afdec6}=-1.68$）。

科克伦-奥克特方法比OLS少报告一个观测值，这说明CO方法没有利用变换后第一次的观测值。这对假设检验中的自由度稍微有点影响。

最后，CO方法估计出的 R-平方也报告出来了，但在这里，它比OLS估计得到的 R-平方要小。但是，实际上这两个 R-平方根本不能进行比较。对OLS来说，R-平方是建立在对没有经过变换的因变量和自变量做回归的基础上的。对CO来说，R-平方通过做变换后的因变量对变换后的自变量的回归而得到的。我们无法断定哪一个 R^2 是正确的，只不过按照习惯将它们报告出来而已。

OLS 和 FGLS 的比较

在科克伦-奥克特或普莱斯-温斯登方法的一些应用中，FGLS 估计量在
391 某些实际的重要方面不同于 OLS 估计量。（例 12.4 不属于这种情况。）这一点虽典型地被解释为可行的 GLS 比 OLS 优越的一个证据，可是事情并没有这么简单。为了说明其中的道理，我们考虑下面一个回归模型：

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

这里的时间序列是平稳的。现在，我们假定大数定律成立，那么，对 β_1 的 OLS 估计就是一致的，如果

$$\mathrm{Cov}(x_t, u_t) = 0 \tag{12.34}$$

表 12.1 **因变量 log（*chnimp*）**

系数	OLS	CO 估计
log(*chempi*)	3.12 (0.48)	2.95 (0.65)
log(*gas*)	0.196 (0.907)	1.05 (0.99)
log(*rtwex*)	0.983 (0.400)	1.14 (0.51)
*befile*6	0.060 (0.261)	-0.016 (0.321)
*affile*6	0.032 (0.264)	-0.033 (0.323)
*afdec*6	-0.565 (0.286)	-0.577 (0.343)
截距项	17.70 (20.05)	37.31 (23.22)
$\hat{\rho}$	—	0.293 (0.084)
观测次数	131	130
R^2	0.305	0.193

早些时候，我们说明了在严格外生性假定下，FGLS 是一致的，而这一
392 假定比式（12.34）更为严格。实际上，可以证明，为了使 FGLS 是一致的，除了式（12.34）以外，最弱的假定是 x_{t-1} 与 x_{t-1} 之和与 u_t 不相关：

$$\mathrm{Cov}(\{x_{t-1} + x_{t+1}\}, u_t) = 0 \tag{12.35}$$

也就是说，FGLS 的一致性要求 u_t 与 x_{t-1}，x_t 及 x_{t+1} 都不相关。

这意味着 OLS 和 FGLS 可能会由于式（12.35）不成立，而产生很不相

同的估计值。在这种情况下，因为 OLS 在式（12.34）下是一致的，而 FGLS 则不一致，所以 OLS 优于 FGLS。如果 x 对 y 有滞后影响，或者 u_t 的变化会引起 x_{t-1} 的变化，那么 FGLS 会产生错误的结果。

由于 OLS 和 FGLS 是不同的估计方法，所以我们绝不期待它们能够产生相同的估计值。在二者产生的 β_j 的估计值相近的情况下，如果有迹象表明序列相关存在，FGLS 方法更可取，因为它的估计量更有效一些，而且 FGLS 统计量至少是渐近正确的。当 OLS 和 FGLS 的估计值有实际差别时，更困难的一个问题出现了：我们很难判断这种差别是否是统计上显著的。豪斯曼（Hausman，1978）提出了一种一般性的方法，但它超出了本书的范围。

OLS 和 FGLS 的一致性和渐近正态性，严重地依赖于时间序列 y_t 和 x_{tj} 的弱相依性质。若把 OLS 或 FGLS 应用于有单位根的过程，就会有一些很奇怪的结果出现。这方面的问题我们将在第 18 章讨论。

例 12.5　静态菲利普斯曲线

表 12.2 列出了例 10.1 中的静态菲利普斯曲线的 OLS 估计结果和 CO 迭代估计的结果。

表 12.2　　**因变量 *inf***

系数	OLS	CO 估计
unem	0.468 (0.298)	-0.665 (0.320)
截距项	1.424 (1.719)	7.580 (2.379)
$\hat{\rho}$	-	0.774 (0.091)
观测次数	49	48
R^2	0.053	0.086

393 比较令人感兴趣的是 $unem_t$ 的系数，用 CO 和 OLS 两种不同的方法得到的估计值差别很大。因为 CO 的估计值更好地反映出了通货膨胀和失业率之间的替代关系，我们倾向于使用 CO 估计。实际上，CO 估计值与把 *inf* 和 *unem* 都进行一阶差分后得到的结果非常接近（见习题 11.11），这是因为，$\hat{\rho}=0.774$ 时在 CO 中使用的准差分与一阶差分非常相近。所以，很可能的情况是，*inf* 和 *unem* 之间不是水平相关的，它们的一阶差分之间有负的关系。

更高阶序列相关的校正

要想校正更高阶的序列相关也是可能的。哈维（Harvey，1990）给出了一种一般性的补救方法。这里，我们只介绍以下AR(2)序列相关的补救方法：

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$$

式中，$\{e_t\}$ 满足AR(1)模型的所有假定。这里的平稳性条件更复杂一些，可以证明，它们包括［见 Harvey（1990）］：

$$\rho_2 > -1, \rho_2 - \rho_1 < 1 \text{ 和 } \rho_1 + \rho_2 < 1$$

例如，如果 $\rho_1 = 0.8$ 且 $\rho_2 = -0.3$，模型就是平稳的；如果 $\rho_1 = 0.7$ 且 $\rho_2 = 0.4$，它就不是平稳的。

我们假定平稳性条件成立，就可以得到能够消除序列相关的变换。在简单的回归模型中，当 $t>2$ 时，很容易做到这一步：

$$y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} = \beta_0(1 - \rho_1 - \rho_2) + \beta_1(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2}) + e_t$$

或

$$\tilde{y}_t = \beta_0(1 - \rho_1 - \rho_2) + \beta_1 \tilde{x}_t + e_t, \quad t = 3, 4, \cdots, n \qquad (12.36)$$

如果知道 ρ_1 和 ρ_2 的值，在得到变换后的变量之后，就会很容易地用OLS估计出这个方程。因为很少能够知道 ρ_1 和 ρ_2，我们只有去估计它们。和往常一样，我们可以利用OLS的残差 $\hat{u}_t$：通过下面的回归求出 $\hat{\rho}_1$ 和 $\hat{\rho}_2$。

$$\hat{u}_t \text{ 对 } \hat{u}_{t-1}, \hat{u}_{t-2} \text{ 回归}, \quad t = 3, \cdots, n$$

［它与回归元是严格外生时检验AR（2）序列相关所采用的回归是一样的。］然后，我们再用 $\hat{\rho}_1$ 和 $\hat{\rho}_2$ 来代替 ρ_1 和 ρ_2 去求得变换后的变量。通过以上方法就可以求出一种可行的GLS估计量。如果我们遇到了多元解释变量，那么每个变量都要进行变换：$\tilde{x}_{tj} = x_{tj} - \hat{\rho}_1 x_{t-1,j} - \hat{\rho}_2 x_{t-2,j}(t>2)$。

对头两次观测的处理还需要一点技巧。可以证明，因变量和每个自变量（包括截距）都应该进行如下变换：

$$\tilde{z}_1 = \{(1 + \rho_2)[(1 - \rho_2)^2 - \rho_1^2]/(1 - \rho_2)\}^{1/2} z_1$$

$$\tilde{z}_2 = (1 - \rho_2^2)^{1/2} z_2 - \{\rho_1(1 - \rho_1^2)^{1/2}/(1 - \rho_2)\} z_1$$

394 式中，z_1 和 z_2 分别为 $t=1$ 和 $t=2$ 时的因变量或自变量。我们在这里就不对变换进行推导了。简单地说，变换后的变量消除了前两次观测之间的序列相关，并且使它们的误差的方差都等于 σ_e^2。

用于时间序列分析的计量经济学软件包，能够较容易地估计AR(q)的误差模型；我们很少需要亲自计算变换后的变量。

12.4 差分和序列相关

在第 11 章，我们把差分作为使积整过程变为弱相依的一种变换而提出来。当我们与高度持久的数据打交道时，就可以从另一个角度看到差分的优越之处。假使我们从简单的模型开始：

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, 2, \cdots \tag{12.37}$$

式中，u_t 遵循AR(1)过程（12.26）。正如在 11.3 节提到的，也正如我们将在第 18 章详细讨论的，当变量 y_t 和 x_t 是一阶积整，即 I（1）的时候，通常的 OLS 推断程序可能非常有误导性。在式（12.37）中的 $\{u_t\}$ 遵循随机游走这种极端的情况下，这个方程没有任何意义，因为 u_t 的方差随着时间 t 上升。而对方程进行差分更符合逻辑：

$$\Delta y_t = \beta_1 \Delta x_t + \Delta u_t, \quad t = 2, \cdots, n \tag{12.38}$$

如果 u_t 遵循随机游走，那么就有 $e_t = \Delta u_t$，它有零均值，当然是个常数，也就是序列不相关的了。这样，我们就可以假定 e_t 和 Δx_t 是不相关的，并可以用 OLS 来估计式（12.38）了。在估计中，我们失去了第一次观测。

即使 u_t 不遵循随机游走，只要 ρ 是正的且比较大，一阶差分往往也是不错的想法：它可以消除大部分的序列相关。当然，式（12.38）和式（12.37）是不同的，但至少我们对式（12.38）中的 OLS 标准误和 t 统计量更有信心一些。方程中有多个解释变量时，结论基本还是一样的。

例 12.6　对利率方程进行差分

在例 10.2 中，我们估计了一个将 3 个月国债利率、通货膨胀和联邦赤字联系起来的方程［见方程（10.15）］。如果我们将方程中得出的残差对它的一期滞后做回归，得到 $\hat{\rho} = 0.530$（0.123），这个结果是统计上大于零的。如果我们差分 $i3_t$，inf 和 def，然后重新计算用于检验AR(1)序列相关的残差，重新做回归，得到 $\hat{\rho} = 0.068$（0.145），因此，不存在序列相关。差分显然消除了序列相关。［此外，有证据表明 $i3$ 包含了一个单位根，inf 也是如此，所以有必要进行差分以得到 I(0) 变量。］

395 正如我们在第 11 章中解释的那样，作出是否进行差分的决定是比较困难的。但是以上内容揭示出了差分的另一个优点，它消除了序列相关。我们将在第 18 章继续讨论这个问题。

问题 12.4 假使在用 OLS 估计一个模型之后，你估计回归式（12.14）中的 ρ 得到 $\hat{\rho}$ =

0.92。针对这一结果，你将采取什么措施？

12.5 在OLS后的序列相关—稳健推断

近些年来，有一种越来越流行的做法，就是用OLS来估计模型，然后再针对相当任意的序列相关（及异方差）形式校正标准误。即使我们知道OLS将是非有效的，但有某些原因促使我们采用这种方法。第一，解释变量可能不是严格外生的。在这种情况下，FGLS连一致都称不上，更不用说有效了。第二，在多数应用FGLS的情况下，往往假定误差遵循AR(1)模型。而OLS估计量对于更为一般形式的序列相关具有稳健性，所以，计算OLS估计量的标准误会更理想一些。

为了说明以上内容，我们来考虑方程（12.4），它是一个简单回归模型的OLS斜率估计量的方差，而这个模型的误差遵循AR(1)过程。我们可以很容易地估计出这个方差，只要把ρ和σ^2的标准的估计值代进去就行了。惟一的问题在于，它也假定了AR(1)模型和同方差性。将这两个假定都放松是可能的。

戴维森和麦金农（Davidson and MacKinnon，1993）给出了兼对异方差和序列相关有稳健性的处理标准误的一般方法。这里，我们只介绍一种简单的方法，用它来计算任何OLS系数的稳健性标准误。

我们这里介绍的处理方法来自于伍德里奇（Wooldridge，1989）。考虑下面的标准多元线性回归模型：

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t,\ t = 1, 2, \cdots, n \tag{12.39}$$

我们曾用OLS估计过它。不妨假设，我们对得到$\hat{\beta}_1$的一个序列相关-稳健性标准误感兴趣，这是不难做到的。把x_{t1}表示为其余的自变量和一个误差项的线性函数：

$$x_{t1} = \delta_0 + \delta_2 x_{t2} + \cdots + \delta_k x_{tk} + r_t \tag{12.40}$$

式中，误差r_t有零均值，且与x_{t2}，x_{t3}，…，x_{tk}都不相关。

那么，可以证明，OLS估计量$\hat{\beta}_1$的渐近方差为

$$\mathrm{Avar}(\hat{\beta}_1) = \left[\sum_{t=1}^{n} E(r_t^2)\right]^{-2} \mathrm{Var}\left(\sum_{t=1}^{n} r_t u_t\right)$$

在无序列相关的假定TS.5′下，$\{a_t = r_t u_t\}$是序列不相关的，于是通常的
396 OLS标准误（同方差条件下的）和异方差-稳健性标准误都是正确的。但如果TS.5′不成立，Avar（$\hat{\beta}_1$）的表达式必须考虑对$t \neq s$时a_t和a_s之间的相关性。实际上，一旦两项之间的时间距离大于几个时期，我们就可以假定它们的相关性为零。别忘了，在弱相依的条件下，相关性必然趋近于零，所以

这个方法是合理的。

仿照内维和韦斯特（Newey and West，1987）的一般性框架，伍德里奇（Wooldridge，1989）证明了 Avar$(\hat{\beta})$ 可以通过下面的步骤加以估计。令“se$(\hat{\beta})$”表示通常的（但不正确的）OLS 标准误，令 $\hat{\sigma}$ 表示用 OLS 估计式（12.39）得到的普通的回归标准误。令 $\hat{r}_t$ 表示通过下面的扩充回归得到的残差：

$$x_{t1} \text{ 对 } x_{t2}, x_{t3}, \cdots, x_{tk} \tag{12.41}$$

（和往常一样，包含一个常数项。）对于某个选定的整数 $g>0$，定义

$$\hat{\nu} = \sum_{t=1}^{n}\hat{a}_t^2 + 2\sum_{h=1}^{g}[1-h/(g+1)]\Big(\sum_{t=h+1}^{n}\hat{a}_t\hat{a}_{t-h}\Big) \tag{12.42}$$

式中

$$\hat{a}_t = \hat{r}_t\hat{u}_t, \quad t = 1,2,\cdots,n$$

它看起来有点复杂，但实际上求起来很容易。在计算标准误时允许有多大程度的序列相关，由式（12.42）中的整数 g 来限定。一旦我们有了 $\hat{\nu}$，$\hat{\beta}_1$ 的**序列相关—稳健性标准误**［serial correlation（SC）-robust standard error］就是

$$\mathrm{se}(\hat{\beta}_1) = [\text{“se}(\hat{\beta}_1)\text{”}/\hat{\sigma}]^2\sqrt{\hat{\nu}} \tag{12.43}$$

换句话说，它就是取 $\hat{\beta}_1$ 的通常的 OLS 标准误除以 $\hat{\sigma}$，然后把所得的结果平方，接着再乘以 $\hat{\nu}$ 的平方根。最后得到的值就可以用来构建 $\hat{\beta}_1$ 的置信区间和 t 统计量。

下面我们就举个例子，看看简单情况下的 $\hat{\nu}$ 是什么样的。当 $g=1$ 时

$$\hat{\nu} = \sum_{t=1}^{n}\hat{a}_t^2 + \sum_{t=2}^{n}\hat{a}_t\hat{a}_{t-1} \tag{12.44}$$

当 $g=2$ 时

$$\hat{\nu} = \sum_{t=1}^{n}\hat{a}_t^2 + (4/3)\Big(\sum_{t=2}^{n}\hat{a}_t\hat{a}_{t-1}\Big) + (2/3)\Big(\sum_{t=3}^{n}\hat{a}_t\hat{a}_{t-2}\Big) \tag{12.45}$$

可见，g 越大，就有越多的项被加进来，以校正序列相关。式（12.42）中的乘数［$1-h/(g+1)$］的意义在于使 $\hat{\nu}$ 实际上是非负的。［内维和韦斯特（Newey and West，1987）证明了这一点。］我们显然需要 $\hat{\nu}\geqslant 0$，因为 $\hat{\nu}$ 是方差的估计值，而且它的平方根出现在式（12.43）中。

397 式（12.43）中的标准误对任意形式的异方差是稳健的。实际上，如果我们去掉式（12.42）中的第二项，那么式（12.43）就变成了我们在第 8 章讨论过的常见的异方差-稳健性标准误了（没有自由度的调整）。

式（12.43）中的标准误所依据的理论是技术性很强的，也有点微妙。别忘了，我们开始时声明我们并不知道具体是什么形式的序列相关。如果事实的确如此，我们怎么选择 g 的值呢？相关理论告诉我们，只要 g 与样本的容量一起增长，式（12.43）对相当任意的序列相关形式都能奏效。它的

意思是，对于更大的样本容量，我们就可以对式（12.42）中的序列相关的大小要求更随便些。最近有很多有关 g 和 n 之间关系的研究，我们就不介绍了。如果是年度数据，选一个较小的 g，比如 $g=1$ 或 $g=2$，对于大多数的序列相关了基本上就够了。如果是季度或月度数据，假定有足够多的数据，g 就应该更大些（比如，对季度数据，就选 $g=4$ 或 8，对月度数据，就选 $g=12$ 或 24)。内维和韦斯特（Newey and West，1987）建议取 $4(n/100)^{2/9}$ 的整数部分作为 g 的值；还有一些人建议取 $n^{1/4}$ 的整数部分。Newey-West建议被运用在计量经济学程序 Eviews®中。举个例子，若 $n=50$（这对第二次世界大战以来的年度数据来说是合理的)，就有 $g=3$。($n^{1/4}$ 的整数部分是 $g=2$。)

下面我们就来总结一下怎样计算 $\hat{\beta}_1$ 的序列相关—稳健性标准误。当然，因为可以把任意一个变量摆到第一个位置上，下面的步骤适用于计算任何斜率系数的标准误。

$\hat{\beta}_1$ 的序列相关—稳健性标准误

(i) 用 OLS 估计方程(12.39)，得到“se($\hat{\beta}_1$)”，$\hat{\sigma}$ 和 OLS 残差 $\{\hat{u}_t: t=1, \cdots, n\}$。

(ii) 通过辅助方程（12.41）计算残差 $\{\hat{r}_t: t=1, \cdots, n\}$，然后构造 $\hat{a}_t=\hat{r}_t\hat{u}_t$（对每个 t)。

(iii) 对于你选定的 g，如式（12.42）中那样计算 $\hat{\nu}$。

(iv) 计算式（12.43）中的se($\hat{\beta}_1$)。

从经验来看，序列相关—稳健性标准误一般要比没有序列相关时的 OLS 标准误大。这是因为，在多数情况下，误差是正序列相关的。但有可能的是，当 $\{u_t\}$ 有很大的序列相关时，一些系数的通常的 OLS 标准误和 SC-稳健性标准误比较相近；$\hat{a}_t=\hat{r}_t\hat{u}_t$ 的样本自相关决定着 $\hat{\beta}_1$ 的稳健性标准误。

采用 SC-稳健性标准误不如采用只适用于异方差的标准误，这有几个理由。第一，大型的横截面比大型的时间序列更为普遍，在前者条件下，异方差-稳健性标准误具有更好的性质。而当有较大程度的序列相关及样本容量较小时（即使这种“小”，大到像 100 这样的值），SC-稳健性标准误表现很差。第二，因为我们必须对方程（12.42）中的 g 作出选择，SC-稳健性标准误的计算不是自动完成的。前面我们提到了，一些计量经济学软件包能够自动选择 g 的值，但你必须遵守这个选择，不能随意变更。

SC-稳健性标准误还不能程序化地计算，还有一个原因，那就是在出现
398 严重的序列相关时，OLS 可能会非常低效，尤其在小样本的情况下。在施行 OLS 并校正了标准误的序列相关后，系数往往是不显著的，或者至少比计算通常的 OLS 标准误的时候得到的结果不那么显著。

当我们怀疑一些解释变量不是严格外生的，以至于像科克伦-奥克特这类方法连一致性都谈不上的时候，OLS 估计后的 SC-稳健性标准误最有用不过了。在有滞后因变量的模型中，使用 SC-稳健性标准误也是正确的，当然，要有足够的理由，使得我们假定这些模型存在序列相关是有道理的。

例 12.7　波多黎各最低工资

我们将计算波多黎各就业方程中最低工资效应的SC-稳健性标准误。在例12.2中，我们发现了比较有力的证据表明AR(1)序列相关的存在。像在那个例子中那样，添加log(*usgnp*)，log(*prgnp*)和一个线性时间趋势作为控制变量。

就业率对最低工资的弹性的OLS估计值是$\hat{\beta}_1 = -0.212\ 3$，标准的OLS标准误是“se($\hat{\beta}_1$)”$=0.040\ 2$。估计的标准误是$\hat{\sigma}=0.032\ 8$。另外，用前面$g=2$时的步骤[见式（12.45)]，得到$\hat{v}=0.000\ 805$。于是，SC-异方差—稳健性标准误为：$\mathrm{se}(\hat{\beta}_1)=[(0.040\ 2/0.032\ 8)^2]\sqrt{0.000\ 805}\approx 0.042\ 6$。有意思的是，这个稳健性标准误只比通常的OLS标准误稍微大一点。稳健性t统计量大约是-4.98，所以估计得到的弹性仍然非常显著。

与此相比较，$\hat{\beta}_1$的迭代CO估计值是$-0.111\ 1$，标准误为$0.044\ 6$。因此，FGLS比OLS估计值更接近于零，于是我们有理白怀疑严格外生性假定不成立。也许，OLS和FGLS估计值的差别可以由抽样误差来解释。究竟是怎样的就很难说了。

在结束本节之前，我要说明一下，要想构造对序列相关—稳健的、F统计量来检验多重假设，是可能做到的。但是我们在这里不打算介绍这么高深的内容。[见Wooldridge(1991b,1995)，Davidson和MacKinnon（1993）的方法。]

12.6　时间序列回归中的异方差性

在第8章，我们讨论了横截面条件下检验和校正异方差的问题。异方差也可能出现在时间序列模型中，它虽不会造成$\hat{\beta}_j$的偏误或不一致，但会使通常的标准误、t统计量和F统计量变得无效，这与横截面条件下的情况一样。

399　在时间序列回归应用中，异方差问题受到的关注不多，因为序列相关误差的问题往往更亟待解决。尽管如此，在时间序列回归中，对异方差进行检验和校正方面的问题还是值得我们简单讨论一下。

既然常见的OLS统计量在假定TS.1′～TS.5′下是渐近生效的，我们的兴趣就在于，同方差假定TS.4′不成立时会出现什么情况呢？假定TS.2′排除了像遗漏变量这样的错误设定以及某种测量误差出现的可能，而TS.5′排除了误差有序列相关的可能。需要强调的是，序列相关的误差所引起的问题，不是检验和校正异方差所能处理的。

异方差—稳健统计量

在研究横截面回归中的异方差时，我们表明了，异方差不影响 OLS 估计量的无偏性和一致性。对时间序列回归也有完全一样的结论，这一点可以通过回顾无偏性所需要的假定（定理 10.1）和一致性所要求的假定（定理 11.1）看出来。

在 8.2 节，我们讨论了如何调整常见的 OLS 标准误、t 统计量和 F 统计量，以适应未知形式的异方差。在假定 TS.1′, TS.2′, TS.3′和 TS.5′下，在时间序列分析中进行同样的调整也会奏效。因此，如果同方差假定是惟一不成立的假定，那么大多数计量经济学软件包还是能够提供有效的推断的。

异方差的检验

有时，我们希望检验时间序列回归中的异方差，尤其是当样本容量相对较小时，使得我们对异方差—稳健统计量的效果也不放心。第 8 章中介绍的检验方法可以直接在这里应用，但要澄清几点，以防止误解。第一，误差 u_t 不能是序列相关的，任何序列相关都会使得异方差检验无效。因此，需要先检验序列相关，而且如果怀疑有异方差，就应该使用异方差-稳健检验来检验序列相关。经过了序列相关的校正，就可以检验异方差了。

第二，考虑下面的用于布劳殊-培甘异方差检验的方程

$$u_t^2 = \delta_0 + \delta_1 x_{t1} + \cdots + \delta_k x_{tk} + v_t \tag{12.46}$$

式中，虚拟假设为$H_0: \delta_0 = \delta_1 = \cdots = \delta_k = 0$。为了使 F 统计量——用 $\hat{u}_t^2$ 代替 u_t^2 作为因变量——有效，我们必须假定误差 $\{v_t\}$ 自身是同方差（就像横截面的情况下那样）和序列不相关的。所有对异方差的标准检验方法，包括 8.3 节的 White 检验，都隐含地作出了以上假定。假定 $\{v_t\}$ 是序列不相关的，就排除了某些特定形式的动态异方差，我们将在后面介绍一些这方面内容。

如果我们在 u_t 中发现了异方差（但 u_t 是序列不相关的），那么就可以使用异方差-稳健检验统计量。还有一个办法，就是使用加权最小二乘法（weighted least squares），就像 8.4 节那样。时间序列情况下的加权最小二乘法与横截面情况下的加权最小二乘法操作起来是一样的。

400 **例 12.8　异方差和有效市场假说**

在例 11.4 中，我们估计了简单的模型：

$$return_t = \beta_0 + \beta_1 return_{t-1} + u_t \tag{12.47}$$

EMH 声称 $\beta_1 = 0$。当我们用 NYSE.RAW 中的数据检验这个假设时，我们得到了 $n = 689$ 时的 $t_{\hat{\beta}_1} = 1.55$。在这种大样本条件下，没有足够的证据能够否定 EMH。而 EMH 指出，给定过去观测到的信息，期望收益是个定值，它并没有谈到条件方差的问题。实际上，异方差的布劳殊-培甘检验需要做 OLS 残差的平方 $\hat{u}_t^2$ 对 $return_{t-1}$ 的回归：

$$\begin{aligned} \hat{u}_t^2 &= 4.66 - 1.104\, return_{t-1} + residual_t \\ &\quad (0.43)\quad (0.201) \\ &\quad n = 689,\ R^2 = 0.042 \end{aligned} \tag{12.48}$$

$return_{t-1}$ 的 t 统计值大约是 -5.5，有力地证明了异方差的存在。因为 $return_{t-1}$ 的系数是负的，我们发现了一个有趣的结论：当以前的收益高时，股票收益率的变动就小一些，反之则相反。这样，我们发现了很多金融研究得出的共同结论：股票收益的期望值不依赖于它的过去值，但是收益的方差却依赖于过去的收益。

问题 12.5 ☞ 如何计算方程（12.47）中异方差的怀特检验？

自回归条件异方差

近些年来，经济学家开始对动态形式的异方差感兴趣。毫无疑问，如果 x_t 包含滞后的因变量，那么，像式（12.46）中那样的异方差就是动态的。但是，动态形式的异方差也会出现在没有动态的回归方程模型中。

为了说明这一点，让我们考虑一个简单的静态回归模型：

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

并假定高斯-马尔科夫假定都成立。这意味着 OLS 估计量是 BLUE。同方差假定的含义是指 $\mathrm{Var}(u_t \mid \mathbf{Z})$ 是恒定的，其中的 $\mathbf{Z}$ 代表 z_t 的 n 个实现。即使 $\mathbf{ZZ}$ 给定时 u_t 的方差是恒定的，异方差还是可能以其他方式出现。恩格尔（Engle，1982）建议，过去的误差（隐含以 $\mathbf{Z}$ 为条件）给定时，应该关注 u_t 的条件方差。恩格尔提出了所谓的**自回归条件异方差**（autoregressive conditional heteroskedasticity，ARCH）模型。一阶 ARCH 模型是

$$\mathrm{E}(u_t^2 \mid u_{t-1}, u_{t-2}, \cdots) = \mathrm{E}(u_t^2 \mid u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2 \tag{12.49}$$

401 在这里，我们隐含了它是以 $\mathbf{Z}$ 为条件的。这个方程仅仅表示出了 $\mathrm{E}(u_t \mid u_{t-1}, u_{t-2}, \cdots) = 0$，即误差序列不相关时 u_t 的条件方差。条件方差必定是正的，所以这个模型只有在 $\alpha_0 > 0$ 且 $\alpha_1 \geq 0$ 时才有意义；如果 $\alpha_1 = 0$，方差方程中就没有动态了。

可以把式（12.49）写得更直观一些：

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + v_t \qquad (12.50)$$

式中，v_t 的期望值（给定 u_{t-1}，u_{t-2}，…）被定义为零。（由于约束条件 $v_t \geqslant -\alpha_0 - \alpha_1 u_{t-1}^2$，$v_t$ 不是独立于过去的 u_t 的。）方程（12.50）看起来很像 u_t^2 的一个自回归模型（因此得名ARCH）。这个方程的平稳性条件是 $\alpha_1 < 1$，和常见的AR(1)一样。如果 $\alpha_1 > 0$，误差的平方就包含（正的）序列相关，即使 u_t 本身不是序列相关的。

式（12.50）对 OLS 有何影响呢？因为我们在开始的时候假定了高斯-马尔科夫假定都成立，OLS 当然是 BLUE。而且，即使 u_t 不是正态分布的，我们也知道，通常的 OLS 检验统计量在假定 TS.1′～TS.5′下也是渐近有效的，而静态模型和有 ARCH 误差的分布滞后模型都满足了这 5 个假定。

如果 OLS 在 ARCH 条件下仍然有比较理想的性质，我们为什么还要关心静态和分布滞后模型中的 ARCH 形式的异方差呢？这里有两个原因。第一，我们有可能得到 β_j 的一致的（但不是无偏的）估计值，它比 OLS 估计值更渐近有效。基于式（12.50）的加权最小二乘法就可以求出这种估计值。在误差 u_t 有条件正态分布的情况下，使用最大似然法也可以。第二，不同领域的经济学家都越来越对条件方差的动态感兴趣。恩格尔最开始将它应用于英联邦通货膨胀的方差，他发现，过去时期的更大规模的误差（更大的 u_{t-1}^2）是与当期的更大的误差方差相联系的。方差经常被用来度量波动，而波动在资产定价理论中又是个关键的内容，所以 ARCH 模型在金融的经验研究中越来越重要。

ARCH 模型也适用于条件均值有动态的情况。例如，假使有因变量 y_t，一个同期外生变量 z_t，及

$$\mathrm{E}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \cdots) = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1}$$

也就是最多有 y 和 z 的一期滞后出现在动态回归中。一般的思路是假定 $\mathrm{Var}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \cdots)$ 为定值，就像我们在第 11 章讨论的那样。但这个方差也可以遵循一个 ARCH 模型：

$$\begin{aligned}\mathrm{Var}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \cdots) &= \mathrm{Var}(u_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \cdots) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2\end{aligned}$$

402 式中，$u_t = y_t - \mathrm{E}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \cdots)$。在第 11 章我们了解到，ARCH 的出现并不会影响 OLS 的一致性，通常的异方差—稳健性标准误和检验统计量都是生效的。（别忘了，它们对任何形式的异方差都是有效的，而 ARCH 不过是一种特殊形式的异方差而已。）

如果你对 ARCH 模型及相关内容感兴趣，请参见 Bollerslev，Chou，Kroner（1992）和 Bollerslev，Engle 和 Nelson（1994），以找到最近的综述。

例 12.9　股票收益的 ARCH

在例 12.8 中，我们看到了，每周股票收益有异方差。这种异方差实际上

可以用式(12.50)中的 ARCH 模型更好地来描述。如果我们计算式(12.47)中的 OLS 残差，再将它们平方，然后做它们对滞后残差平方的回归，得到

$$\hat{u}_t^2 = 2.95 + 0.337\hat{u}_{t-1}^2 + residual_t \qquad (12.51)$$
$$(0.44) \quad (0.036)$$
$$n = 688, R^2 = 0.114$$

$\hat{u}_{t-1}^2$的 t 统计量大于9，表明有很强的 ARCH。与我们在前面得出的结论一样，时间 $t-1$ 的较大的误差意味着今天股票收益的更大的方差。

重要的是要看到，虽然 OLS 残差的平方是自相关的，但 OLS 残差本身却不是自相关的（这与 EMH 是一致的）。做 $\hat{u}_t$ 对 $\hat{u}_{t-1}$的回归得到 $\hat{\rho}$，$t_{\hat{\rho}} = 0.038$。

回归模型中的异方差和序列相关

在回归模型中，同时出现异方差和序列相关是有可能的。如果没有把握，我们总是可以使用 OLS 方法并计算一个全面稳健性标准误，就像 12.5 节描述的那样。

很多时间序列中的相关性问题被当做最重要的问题，因为它往往比异方差对标准误和估计量的效率有更大的影响。在 12.2 节我们已经认识到，要想得到对任何形式的异方差的序列相关都稳健的检验并不难。如果我们用这种检验来探测序列相关，可以采用科克伦-奥克特变换［见方程（12.32)］，然后再使用变换后的方程中的异方差-稳健性标准误和检验统计量。或者，我们也可以在方程（12.32）中使用布劳殊-培甘或怀特检验来进行异方差检验。

另外还有一种方法：我们可以分别建立异方差和序列相关的模型，再通过加权最小二乘法AR(1)组合程序来校正异方差和序列相关。下面我们来看一个特殊一点的模型：

408

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \\ u_t &= \sqrt{h_t}v_t \qquad (12.52) \\ v_t &= \rho v_{t-1} + e_t, |\rho| < 1 \end{aligned}$$

其中的解释变量 **X** 在所有的时间 t 都是独立于e_t 的，而 h_t 是x_{tj}的一个函数。过程 $\{e_t\}$ 的均值为0，方差为 σ_e^2，且是序列不相关的。因此，$\{v_t\}$ 符合稳定的AR(1)过程的所有条件。略去解释变量的条件后，得到

$$\mathrm{Var}(u_t) = \sigma_v^2 h_t$$

式中，$\sigma_v^2 = \sigma_e^2/(1-\rho^2)$。但 $v_t = u_t/\sqrt{h_t}$是同方差的，且遵循一个稳定的AR(1)模型。因此，变换后的方程

$$y_t/\sqrt{h_t} = \beta_0(1/\sqrt{h_t}) + \beta_1(x_{t1}/\sqrt{h_t}) + \cdots + \beta_k(x_{tk}/\sqrt{h_t}) + v_t \tag{12.53}$$

有AR(1)形式的误差。现在，如果我们在头脑中有一个特定形式的异方差——就是说，知道 h_t——我们就可以用标准的 CO 或 PW 方法来估计式(12.52)。

在多数情况下，我们必须首先估计 h_t。下面的方法将8.4 节讲到的加权最小二乘法和12.3 节讲到的AR(1)序列相关结合在一起。

异方差和AR(1)序列相关的可行的 GLS

(i) 用 OLS 估计式（12.52）并记住残差。

(ii) 做 log（$\hat{u}_t$）对 x_{t1}，…，x_{tk}（或对 $\hat{y}_t$，$\hat{y}_t^2$）的回归，得到拟合值，记为 $\hat{g}_t$。

(iii) 求出 h_t 的估计值：$\hat{h}_t = \exp$（$\hat{g}_t$）。

(iv) 用标准的科克伦 奥克特或普莱斯 温斯登方法估计下面变型后的方程

$$\hat{h}_t^{1/2}y_t = \hat{h}_t^{-1/2}\beta_0 + \beta_1\hat{h}_t^{1/2}x_{t1} + \cdots + \beta_k\hat{h}_t^{1/2}x_{tk} + \text{误差项} \qquad (12.54)。$$

这些可行的 GLS 估计量是渐近有效的。更为重要的是，所有通过 CO 或 PW 方法得出的标准误和检验统计量都是渐近有效的。

▶ 小结

我们讨论了多元回归模型中的误差的序列相关这一重要问题。相邻的误差之间的正的序列相关是较为普遍的，特别是在静态和有限分布滞后模型中。它导致通常的 OLS 标准误和统计量有误导性。(尽管 $\hat{\beta}_j$ 仍然可能是无偏的，或者至少是一致的。) 一般来讲，OLS 标准误会低估参数估计值的真实不确定性。

最常见的序列相关模型是AR(1)模型。把它作为研究的起点，用 OLS 残差就可以容易地进行AR(1)序列相关的检验。假定回归元是严格外生的，并且同方差条件成立，就可以通过做 OLS 残差对滞后残差的回归得到渐近生效的 t 统计量。使检验对异方差有稳健性并不难。在经典线性模型假定 404 下，可以使用德宾-沃森统计量，但它可能会得出没有结论的结果，而且，它并不比 t 检验好多少。

对于有滞后因变量或其他非严格外生回归元的模型来说，$\hat{u}_{t-1}$的标准的 t 检验还是有效的，只要所有的自变量都与 $\hat{u}_{t-1}$ 一道被当做回归元加进模型中来。我们还可以使用 F 统计量或 LM 统计量来检验更高阶的序列相关。

在回归元是严格外生的模型中，我们可以使用一种可行的 GLS 程序——科克伦-奥克特方法或普莱斯 温斯登方法——来校正AR(1)序列相关。

它给出了不同于OLS估计值的估计值：FGLS估计值是通过对准差分变量进行OLS估计得到的。所有从变换后的方程中得到的常见的检验统计量都是渐近生效的。几乎所有的回归软件包都具有估计带有AR(1)误差的模型的功能。

另一种补救序列相关的办法，特别是在严格外生性假定不成立的时候，是用OLS去计算序列相关—稳健性标准误（对异方差它也是稳健的）。很多回归软件包采用的是内维和韦斯特（Newey and West，1987）提出的方法；用标准的回归软件包每次只求出一个标准误也是可能的。

最后，我们讨论了时间序列模型中的异方差性的一些特点。和横截面分析中一样，最重要的一种异方差是取决于解释变量的异方差；它决定着常见的OLS估计量是否生效。第8章中介绍的布劳殊-培甘和怀特检验可以直接应用，只要误差不是序列相关的就行了。近些年来，经济学家——特别是研究金融市场的经济学家——对动态形式的异方差越来越感兴趣。ARCH模型就是个明显的例子。

关键术语

自回归条件异方差性（ARCH）
布劳殊-戈弗雷检验
科克伦-奥克特（CO）估计
德宾-沃森（DW）统计量
可行GLS（FGLS）
普莱斯-温斯登（PW）估计
准差分数据
序列相关—稳健标准误
加权最小二乘法

习　题

12.1　当回归模型中的误差有AR(1)序列相关时，为什么OLS标准误倾向于低估$\hat{\beta}_j$的抽样方差？OLS标准误总是过小吗？

12.2　解释下面的表述有何不当之处："科克伦-奥克特方法和普莱斯-温斯登方法都是用来求OLS估计值的真实标准误的。"

405　12.3　在例10.6中，我们估计了一个费尔模型的变形，来预测美国总统选举的结果。

(i) 对于这个方程中的误差项序列不相关你有何见解？［提示：总统选举多长时间进行一次？］

(ii) 当你做式（10.23）的OLS残差对滞后残差的回归时，得到$\hat{\rho}=-0.068$和$\mathrm{se}(\hat{\rho})=0.240$。你对$u_t$中的序列相关有何结论？

(iii) 在检验序列相关时，较小的样本容量会令你不放心吗？

12.4 判断对或错："如果回归模型的误差包含 ARCH，它们一定是序列相关的。"

12.5 （i）在习题 10.11 的工业区事件研究中，OLS 残差对滞后残差的回归给出 $\hat{\rho}=0.841$ 和 se（$\hat{\rho}$）$=0.053$。这对于 OLS 来说意味着什么？

（ii）如果你想使用 OLS，而且还想求出 EZ 系数的一个真实的标准误，你将怎样做到？

12.6 在例 12.8 中，我们发现了方程（12.47）的误差中存在异方差的证据。那么，我们就来计算异方差-稳健性标准误［在［·］里］及通常的标准误：

$$\begin{aligned}\widehat{return}_t &= 0.180 + 0.059\, return_{t-1}\\ &\quad (0.081)\ \ (0.038)\\ &\quad [0.085]\ \ [0.069]\\ &\quad n=689,\ R^2=0.003\,5,\ \bar{R}^2=0.002\,0\end{aligned}$$

使用异方差—稳健 t 统计量对 $return_{t-1}$ 的显著性有何影响？

计算机习题

12.7 在例 11.6 中，我们估计了一个一阶差分形式的有限分布滞后模型：

$$\Delta gfr_t = \gamma_0 + \delta_0 \Delta pe_t + \delta_1 \Delta pe_{t-1} + \delta_2 \Delta pe_{t-2} + u_t$$

利用 FERTIL3.RAW 中的数据来检验误差中是否存在AR(1)序列相关。

12.8 （i）利用 WAGEPRC.RAW 中的数据，估计习题 11.5 中的分布滞后模型。用回归式（12.14）来检验AR(1)序列相关。

（ii）用迭代的科克伦-奥克特方法重新估计这个模型。新的长期倾向的估计值是多少？

（iii）用迭代的 CO 求出 LRP 的标准误是多少。（这要求你估计一个修改过的方程。）估计到的 LRP 在 5% 的水平上是统计上异于 1 的吗？

406 12.9 （i）在习题 11.13 的（i）部分，你估计了存货的加速模型。检验这个方程中的 AR（1）序列相关。

（ii）如果你发现存在着序列相关，请用科克伦-奥克特方法重新估计这个方程，并将所得结果与以前的结果进行比较。

12.10 （i）利用 NYSE.RAW 中的数据估计方程（12.48）。令 $\hat{h}_t$ 表示这个方程中的拟合值（条件方差的估计值）。多少个 $\hat{h}_t$ 是负的？

（ii）在式（12.48）中加进 $return_{t-1}^2$，然后再计算拟合值 $\hat{h}_t$。哪些 $\hat{h}_t$ 是负的？

（iii）利用（ii）部分得到的 $\hat{h}_t$ 估计式（12.47），要求采用加权最小二乘法（像在 8.4 节中那样）。将 β_1 的估计值与方程（11.16）中的相应结果比较一下。

（iv）现在用 WLS 估计方程（12.47），并用式（12.51）中估计得到的

ARCH 模型求出 $\hat{h}_t$。这时，你的结果与（iii）中的是否相同？

12.11　考虑例 10.6 中的那个 Fair 模型。现在，我们不去预测民主党在两党选举中的得票比例，而去估计一个表示民主党是否获胜的线性概率模型。

（i）用虚拟变量 *demwins* 来代替式（10.23）中的 *demvote*，并用标准的格式报告出结果。哪些因素影响获胜的概率？请用到 1992 年为止的数据。

（ii）有多少个拟合值小于 0？有多少个拟合值大于 1？

（iii）采用下面的预测规则：如果 $demwins > 0.5$，你就可以预言民主党会获胜；否则，共和党将获胜。那么，在这 20 次选举中，共有多少次能被这个模型正确地预测其结果？

（iv）加进解释变量在 1996 年的值。预测克林顿赢得这次选举的可能性有多大。事实上，克林顿获胜了，你的预测结果是否与事实相符？

（v）对误差中的AR(1)序列相关，做异方差—稳健 t 检验。你有何发现？

（vi）求出（i）部分中的估计值的异方差—稳健性标准误。t 统计量有什么明显的变化吗？

12.12　（i）在习题 10.13 中，你估计了消费增长和可支配收入之间的一种简单的关系。检验这个方程中的 AR（1）序列相关（用 CONSUMP.RAW）。

（ii）在习题 11.14 中，你通过消费的增长对它的一期滞后的回归，检验了持久收入假说。在做这个回归之后，再通过残差平方对 gc_{t-1} 和 gc_{t-1}^2 的回归来检验异方差，你有何结论？

经济科学译丛·计量经济学导论·现代观点　经济科学译丛·计量经济学导论·现代观点　经济科学译丛·计量

第3篇

高深专题讨论

407 我们现在转到一些专门的问题上来，这些专题不一定在一个一学期的入门课程中有所介绍，其中的一些还需要用到比第1,2两篇的多元回归分析更多的数学技巧。第13章阐述怎样把多元回归应用到独立混合的几个横截面上，这里涉及的问题，除了我们可以通过引进一些时间虚拟变量得以研究关系式如何随时间而变化外，非常类似于标准的横截面分析。我们还将阐明怎样能在回归的框架中，分析综列数据集。第14章讨论较高深然而经常见于应用研究的综列数据方法。

第15、16两章考虑内生解释变量的问题。在第15章中，我们介绍工具变量法作为解决遗漏变量问题和测量误差问题的一种方法。二阶最小二乘法常用于经验经济学而且是估计联立方程模型所不可缺少的，这就是第16章要讨论的一个专题。

第17章包括一些用于横截面分析的、相当高深的专题，如限值因变量模型和样本选择偏误的纠正方法。第18章则转到另一个方向，介绍了一些在动态关系式估计中行之有效的时间序列计量经济学的新近进展。

第19章对于必须写出学期论文或其他应用社会科学论文的学生应有所帮助，该章对如何选题、收集并分析数据以及论文写作，都提出了建议。

第 13 章　跨时横截面的混合，简单综列数据方法

408　直到现在，我们所讨论的多元回归分析，要么使用纯粹的横截面数据，要么使用纯粹的时间序列数据。虽然这两种情形都常见于实际应用，但在经验（实证）研究中也越来越多地用到兼有横截面和时间序列因次（维数）的数据集。事实上，兼有横截面和时间序列两个方面的数据，常能给重要的政策问题研究带来曙光。我们将在本章中展示几个例子。

本章将分析两种数据集。一种是**独立混合横截面**（independently pooled cross section），它是在不同时点（经常但并不一定是不同的年份）从一个大总体里进行随机抽样的结果。例如，我们每年从美国在职的工作人员总体里抽取一个关于小时工资、学历、工作经验等等的随机样本。或者，我们每隔一年就对某大都市区出售的住房抽取一个关于售价、面积的平方公尺、浴室间数的随机样本。从统计学的观点看，这些数据集有一个重要的特点：它们都是由独立抽取的观测值构成的。这也是我们做横截面分析时的主要依据：这至少排除了在不同的观测中误差项有相关关系。

一个独立混合横截面和单一个随机样本的差异在于，在不同时点上对总体进行抽样很可能导致观测点（即观测结果）不是**同分布**（identically distributed）的情形。比如说，随着时间的流逝，大多数国家的工资和学历的分布已经改变。我们将会看到，这实际上是一个容易对付的问题，即可在多元回归模型中，容许截距以至在某些情形中还容许斜率随时间而改变。在 13.1 节中，我

们就考虑这样的模型。在 13.2 节中，我们将讨论如何把不同时间的横截面混合起来，用于评价政策的改变。

另一种是**综列数据**（panel data）集。它虽然兼有横截面和时间序列的因次，但在一些重要方面却不同于独立混合横截面，如收集综列数据——有时又称**纵横数据**（longitudinal data）——我们要在不同时间跟踪（或试图跟踪）相同的一些个人、家庭、厂家、城市、州或别的什么单元。例如，在一个时点上，从某总体中随机地收集了一些人的个人工资、工作小时、学历和其他因素的一个综
409 列数据集，那么，在以后的若干个时点上，要对同样的这些人反复采访，以便得到同样一群人在不同年份里的工资、工作小时、学历等数据。

从校区、城市、县、州和乡村收集综列数据是相当容易的，从而大大推动了用综列数据集去做政策分析。就综列数据的计量经济而言，我们可不能假定不同时点的观测值是独立地分布的。例如，影响着某人 1990 年工资收入的那些不可观测的因素仍将影响该人在 1991 年的工资；影响着某城市 1985 年犯罪率的不可观测因素仍将影响该城市在 1990 的犯罪率。因此，还须研究出可用以分析综列数据的特殊模型和方法。在 13.3 节、13.4 节和 13.5 节里，我们将描述简单的差分法，以消除所研究单元的不随时间而变的那些观测不到的特征。由于综列数据方法比较深奥，我们将主要凭直观来描述估计程序中的统计性质，而把推导细节留作章末附录。在第 14 章讨论更复杂的综列数据方法时，我们将采取同样的策略。

13.1 跨时独立横截面的混合

许多关于个人、家庭和厂商的调查，每隔一段时间，常常是每隔一年，重复进行一次。一个例子是当前人口调查（current population survey, CPS），它每年都对家庭随机地抽查一次，例如，参看含有 1978—1985 年调查数据的 CPS78-85.RAW。如果每个时期都抽取一个随机样本，那么把所得到的随机样本合并起来就给出一个独立混合横截面。

使用独立混合横截面的一个理由是要加大样本容量，把在不同时点从同一总体中抽取的多个随机样本混合起来使用，可以获取更精密的估计量和更具功效的检验统计量，仅当因变量和某些自变量保持着不随时间而变的关系时，混合才会是有用的。

如在引言中所提到的，使用混合横截面只会带来少量的统计复杂性。典型地说，总体在不同时期会有不同的分布。为了反映这一事实，我们允许截距在不同时期，通常是不同年份，有不同的值。通过虚拟变量的引进比如除掉一年外每年加一个虚拟变量，并按照通常方式把最早的一年选做基年，就可容易地达到这一目的。此外，误差方差还可能随时间而变，这正是我们以后要讨论的一些内容。

有时，年虚拟变量的系数变化模式本身就是我们所感兴趣的。例如，人口

学家也许对下述问题感兴趣，把教育加以控制后，问 35 岁以上妇女的生育模式在 1972—1984 年间有无变化。下面的例子说明怎样利用带有年虚拟变量的多元回归分析，可以回答这一问题。

例 13.1　不同时期的妇女生育率

FERTIL1.RAW 中的数据库，类似于桑德(Sander,1994)所用的来源于民意研究中心(National Opinion Research Center)1972—1984 年间(包括 1972 年 410 和 1984 年)的双年社会总调查(General Social Survey)。我们利用其中的数据来估计一个用以解释一个妇女生育小孩总数(*kids*)的模型。

一个令人感兴趣的问题是：在控制了其他可观测因素之后，这段时间里的生育率出现过什么变化？我们所控制的因素是受教育的年数、年龄、种族、16 岁时的生活地区、16 岁时的生活环境。估计结果由表 13.1 给出。

基年是 1972 年。年虚拟变量的系数表明在 80 年代早期生育率有一个明显下落。例如，1982 年的系数意味着在保持教育、年龄和其他因素不变的情况下，1982 年和 1972 年相比，一位妇女平均少生育 3.52 个孩子，或者说大约少 3 个半孩子。这是一个很大的下跌：若保持教育、年龄和其他因素不变，在 1982 年每 100 个妇女将预料比 1972 年可比的妇女少生育约 52 个小孩。因为我们控制了教育，这一下跌就和因平均教育水平的提高而导致的生育率下降没有联系。(1972 年的平均教育年龄为 12.2，而 1984 年为 13.3。)*y*82 和 *y*84 代表解释变量所不能解释的生育率下降。

既然 1982 年和 1984 年虚拟变量的系数都是个别地非常显著的，那么把多个年虚拟变量看做一组变量，也是联合地非常显著的，就没有什么奇怪的了。没有年虚拟变量的回归，R-平方是 0.101 9，并从而算出 $F_{61\ 111} = 5.87$ 和 p 值 $=0$。

多受教育的妇女有较少的小孩，并且估计值是非常显著的。在其他条件不变的情况下，100 名受大学教育的妇女和 100 名仅受高中教育的妇女相比，生育的小孩要少 51 个：0.128(4)=0.512。年龄对生育有抑制的作用。(二次式的转折点在年龄=46 处。到了这个年龄，大多数妇女已停止生育小孩。)

最后，所估计的方程也许隐藏着误差项的异方差性。可利用第 8 章的方法来处理这个问题。但这里有一个令人感兴趣的差别：误差方差即使不随教育、年龄、肤色而变，还可能随时间而变。然而，异方差—稳健标准误及其检验统计量是用得上的。通过平方 OLS 残差对表 13.1 所有的自变量(包括年虚拟变量)回归，就能作出 Bneusch-Pagan 检验。(至于 White 统计量这个特殊情况，和平常一样，还要把拟合值 $\widehat{kids}$ 及其平方用做自变量。)加权最小二乘程序应能解决误差可能随时间而变的问题。在 8.4 节所讲的程序中，还要把年虚拟变量放到方程(8.32)中去。

问题 13.1

在阅读表 13.1 时，有人声称如果表中的其他各项均保持不变，则可以预期，一名黑人妇女要比一名非黑人妇女多生育一个孩子，你同意吗？

还可通过一个年虚拟变量和某些主要解释变量之间的交互作用，来考察这些变量的作用是否在某一时期发生了变化。下一个例子分析教育回报和性别差异（歧视）是否从 1978—1985 年已发生变化。

411 **表 13.1** 妇女生育的决定因素

因变量：*kids*

自变量	系数	标准误
educ	-0.128	0.018
age	0.532	0.138
age^2	-0.005 8	0.001 6
black	1.076	0.174
east	0.217	0.133
northcen	0.363	0.121
west	0.198	0.167
farm	-0.053	0.147
othrural	-0.163	0.175
town	0.084	0.124
smcity	0.212	0.160
y74	0.268	0.173
y76	-0.097	-0.179
y78	-0.069	-0.182
y80	-0.071	-0.183
y82	-0.522	0.172
y84	-0.545	0.175
常数项	-7.742	3.052

$n = 1\ 129$

$R^2 = 0.129\ 5$

$R^2 = 0.116\ 2$

412 例 13.2 教育回报和工资中性别差异的变化

跨越 1978 年（基年）和 1985 年的一个混合对数（工资）方程（其中工资以每小时工资计）是

$$\log(wage) = \beta_0 + \delta_0 y85 + \beta_1 educ + \delta_1 y85 \cdot educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 union + \beta_5 female + \delta_5 y85 \cdot female + u \qquad (13.1)$$

式中大多数解释变量我们都应已熟悉。变量 *union*（工会）是一个虚拟变量；如果某人是工会会员，它就等于 1，否则等于 0。变量 *y*85 也是一个虚拟变量；如果观测值来自 1985 年就等于 1，如果来自 1978 年就等于 0。在 1978 年的样本中有 550 人，而在 1985 年则是含有不同的 534 人的另一组人。

1978 年的截距是 β_0，而 1985 年的截距是 $\beta_0 + \delta_0$。1978 年的教育回报是 β_1，而 1985 年的教育回报是 $\beta_1 + \delta_1$。因此，δ_1 度量着多受一年教育的回报，经过 7 年的时间已发生了什么变化。最后，男女对数工资的差别在 1978 年是 β_5；在 1985 年是 $\beta_5 + \delta_5$。于是，通过检验 H_0：$\delta_5 = 0$ 就能检验性别差异在这 7 年里没有变化的虚拟假设。表示性别差异已减少的对立假设则是 H_1：$\delta_5 > 0$。（性别差异意指歧视妇女——译者注）。为简单起见，方程 (13.1) 假定了工作经验和工会会员资格在两个时期里对工资都有同样的影响。

在我们分析估计的结果之前，仍有一个问题需要澄清，就是每小时工资是以名义（或当时）美元计算的。因为名义工资可以仅因通货膨胀而增加，而我们真正感兴趣的却是每个解释变量对真实工资的影响，那么，假定我们决定用 1978 年美元来度量工资，这就要求我们把 1985 年工资平缩为 1978 年美元。（利用 1997 年总统经济报告中的消费者物价指数，平缩因子就是 107.6/65.2≈1.65。）虽然用 1.65 去除每人在 1985 年的工资是容易的，但我们将发现，如果回归中含有一个 1985 年变量，并且用对数工资（而不是工资）作为因变量，就没有必要这样做。取对数形式，用真实工资也好，名义工资也好，只会影响年虚拟变量 *y*85。为了看到这点，令 *P*85 为 1985 年工资的平缩因子（如果用 CPI 它就是 1.65）。然后取 1985 年样本中第 *i* 人的真实工资的时数，就是

$$\log(wage_i / P85) = \log(wage_i) - \log(P85)$$

式中，尽管每人的工资不同，而 log（*P*85）却是相同的，因此 log（*P*85）将被吸收到 1985 年的截距中来。（如果，比方说，对不同地区的人，使用不同的物价指数，这个结论便要改变。）基本要点是，为了研究教育回报和性别差异曾经发生了怎样的变化，我们不需要在方程式（13.1）中把名义工资转换成真实工资。习题 13.8 要求你对现在这个例子证实这一点。

如果我们忘记对 1978 年和 1985 年考虑不同的截距，那么使用名义工资将会产生严重的误导性结果。如果使用工资而不是对数工资，那么记住使用真实工资并且引入一个年虚拟变量是重要的。

413 无论我们用美元价值去表示因变量或自变量，以上讨论一般地说都是对

的。只要美元数额是以对数形式出现，并且对所有时期都采用了虚拟变量（当然，基期除外），总物价平缩因子的使用只会影响那些截距，而不会改变任何一个斜率估计值。

现在我们用 CPS78-85.RAW 的数据来估计方程式

$$
\begin{aligned}
\log(\widehat{wage}) = {} & \underset{(0.093)}{0.459} + \underset{(0.123)}{0.118}\,y85 + \underset{(0.0067)}{0.0747}\,educ + \underset{(0.0094)}{0.0185}\,y85\cdot educ \\
& + \underset{(0.0036)}{0.0296}\,exper - \underset{(0.00008)}{0.00040}\,exper^2 + \underset{(0.030)}{0.202}\,union \qquad (13.2) \\
& - \underset{(0.037)}{0.317}\,female + \underset{(0.051)}{0.085}\,y85\cdot female \\
& n = 1\,084, R^2 = 0.426, R^2 = 0.422
\end{aligned}
$$

1978 年的教育回报估计约为 7.5%；1985 年的教育回报约为 9.35%，即高出了 1.85 个百分点。由于交互作用项的 t 统计量为 0.018 5/0.009 4≈1.97，教育回报的差异在双侧对立假设下是在 5% 水平上统计显著的。

性别差异怎样呢？在 1978 年，其他条件一样的妇女工资比男性工资约少 31.7%（更准确的估计是 27.2%）。到 1985 年，log（$wage$）的差异是 −0.317+0.085= −0.232。因此，性别差异看来从 1978 年到 1985 年降低了的 8.5 个百分点。交互作用项的 t 统计量约为 1.67，这意味着相对于正的单侧对立假设来说，它是在 5% 水平上显著的。

如果我们在方程（13.2）中考虑所有的自变量与 $y85$ 的交互作用，又会出现什么情况？这等同于估计两个不同的方程，一个对 1978 年和另一个对 1985 年。有时这样做是合适的。例如，在第 7 章，我们曾讨论克鲁格（Krueger，1993）的一项研究。他估计了工作中使用电脑的回报。他分别估计了两个方程，一个使用 1984 年 CPS 数据，另一个使用 1989CPS 数据。通过对不同时期教育回报的变化以及对是否使用电脑有所控制进行了比较，他估计出，在这 5 年期间，所观测到的教育回报的增加，有 1/3～1/2 可归功于电脑的使用［见 Knueger（1993）中的表Ⅷ和Ⅸ］。

对跨越时间的结构性变化做邹至庄检验

在第 7 章，我们讨论过，邹至庄检验——不外是一种 F 检验——怎样能用来决定两组数据之间的多元回归函数有无差别，我们同样可把这种检验用于不同的两个时期。检验的一种形式是，把来自混合估计的残差平方和看做受约束的 SSR；不受约束的 SSR 则是对两个时期分别估计而得的两个 SSR
414 之和。计算这些统计量的具体步骤——如 7.4 节所讲述的。还可找到异方差—稳健的估计方法（见 8.2 节）。

例 13.2 给出对两个时期的计算邹至庄检验统计量的另一方法：先将每

一变量对两个年虚拟变量之一形成交互作用，再检验这个年虚拟变量和全部交互作用项是否联合地显著。由于回归模型中的截距常随时间而变（比方说，在住房价格一例中，起因于通货膨胀），这个内涵较充实的邹至庄检验就能识破是否有这种变化。通常人们更感兴趣的是，设置一个截距差异，然后检验某些斜率系数是否随时间而变（像我们在例 13.2 所做的那样）。

还可对多于两个时期进行邹至庄检验，但计算上可能繁琐。通常是在考虑截距的差异后，通过所感兴趣的变量和年虚拟变量的交互作用，以检验某些斜率系数的恒定性。（参看习题 13.7 和习题 13.8。）

13.2 利用混合横截面做政策分析

混合横截面对于评价某一事件或政策的影响可能非常有用。下面的事件研究案例表明，两个横截面数据集，一个收集于事件发生之前，另一个收集在事件发生之后，怎样可以用来判断该事件的经济效果。

例 13.3　垃圾焚化炉的区位对住房价格的影响

基尔和麦克莱恩（Kiel and McClain，1995）曾研究马萨诸塞州北安德沃市的一个新建的垃圾焚化炉对住房价值的影响。他们利用了多年的数据并作了相当复杂的计量经济分析，我们将只利用两年的数据和一些简化模型，但我们的分析和他们的相类似。

1978 年开始传说要在北安德沃市兴建一座垃圾焚化炉，而于 1981 年动工了，人们预料动工后不久焚化炉便会投入运转；事实上 1985 年才开始运转。我们将利用 1978 年住房出售的价格数据和 1981 年售价的另一个样本数据。我们的假设是靠近焚化炉的房价要比远离焚化炉的房价低。（指维持假设——译者注）。

为了明确，如果房子位于焚化炉 3 公里以内，我们就说它靠近。［但我们要求你像基尔和麦克莱恩（Kiel and McClain，1995）那样，使用房子到焚化炉的实际距离。］我们先来看看远近对房价的以美元计的影响，这就要求我们用不变美元来度量价格。我们一律用波士顿住房价格指数按 1978 年美元计算房价，令 *rprice* 为真实住房价格。

一位天真的分析者会仅仅使用 1981 年的数据并估计一个非常简单的模型：

$$rprice = \gamma_0 + \gamma_1 nearinc + u \tag{13.3}$$

415 式中，*nearinc* 为住房靠近焚化炉时等于 1，否则等于 0 的一个二值变量。用 KIELMC.RAW 数据估计方程，得

$$\widehat{rprice} = \underset{(3\,093.0)}{101\,307.5} - \underset{(5\,827.71)}{30\,688.27}\,nearinc$$

$$n = 142, R^2 = 0.165 \tag{13.4}$$

因为这是一个仅对单个虚拟变量的简单回归，所以截距就是靠近焚化炉的住房的平均售价，而 *nearinc* 的系数则代表靠近焚化炉与远离焚化炉的住房平均售价之差。估计结果表明前者的平均售价比后者的要低 30 688.27 美元。t 统计量的绝对值大于 5；从而我们可以强有力地拒绝靠近和远离焚化炉的住房有相同（原书中误为不相同——译者注）价值的假设（指虚拟假设——译者注）。

不幸的是，方程（13.4）并不意味着焚化炉的坐落是造成较低房价的原因。其实，如果对 1978 年（在尚无焚化炉传说之前）做同样的回归，我们得到

$$\begin{aligned} \hat{rprice} &= 82\ 517.23 - 18\ 824.37\,nearinc \\ &\quad (2\ 653.79)\quad (5\ 827.71) \\ &\quad n = 179, R^2 = 0.082 \end{aligned} \tag{13.5}$$

因此，即使在设有任何关于焚化炉的传说之前，靠近其坐落的平均房价就比远离它的平均房价（82 517.23 美元）低了 18 824.37 美元；而且这一差额也是统计上显著的。这正符合焚化炉本来就要建造在房价较低地带的观点。

这样一来，我们怎样能说新建一个焚化炉会压低房价呢？关键在于看到 *nearinc* 的系数 1978—1981 年间的变化，1981 年的平均房价差异比 1978 年的要大得多（30 688.27 美元比 18 824.37 美元），即使把差异折算成不靠近焚化炉的平均房价的百分比，也是不小的，*nearinc* 的两个系数之差是

$$\hat{\delta}_1 = -30\ 688.27 - (-18\ 824.37) = -11\ 863.9$$

这是我们对焚化炉对其附近房价的影响所做的估计。在经验经济学中，$\hat{\delta}_1$ 曾被称为差异中的差分估计量（difference-in-differences estimator），因为它可表示为

$$\hat{\delta}_1(\overline{rprice}_{81,nr} - \overline{rprice}_{81,fr}) - (\overline{rprice}_{78,nr} - \overline{rprice}_{78,fr}) \tag{13.6}$$

式中，nr 代表靠近焚化炉坐落；而 fr 代表远离焚化炉坐落。换句话说，$\hat{\delta}_1$ 是两个地区的平均房价价差在不同时间上的差分。

为了检验 δ_1 是否显著异于零，我们需要通过回归分析，求出它的标准误。其实，$\hat{\delta}_1$ 可通过估计

$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 \cdot nearinc + u \tag{13.7}$$

416 而获得，估计时用到两个年份的混合数据。截距 β_0 代表 1978 年不靠近焚化炉的一所住房的平均价格。参数 δ_0 概括了北安德沃市的全部住房价值 1978—1981 年的变化。［比较一下方程（13.4）和方程（13.5），即表明相对于波士顿住房价格指数而言，北安德沃的房价在这一时期急剧上升。］*nearinc* 的系数 β_1 度量着与焚化炉的出现无关的区位效应：如同我们在方程（13.5）中看到的，即便在 1978 年，靠近焚化炉位置的住房售价就低于较为远离焚化炉的住宅。

我们关注的参数是交互作用项 $y81 \cdot nearinc$ 的系数 δ_1：δ_1 是房价因新建的焚化炉而下跌的尺度，如果住房不论是靠近抑或远离炉址都不因其他理由而按不同比率升值的话。

方程（13.7）的估计值见表 13.2 中的第（1）列。

表 13.2 **因变量：*rprice***

自变量	(1)	(2)	(3)
常数项	82 517.23 (2 726.91)	89 116.54 (2 406.05)	13 807.67 (11 166.59)
$y81$	18 790.29 (4 050.07)	21 321.04 (3 443.63)	13 928.48 (2 798.75)
$nearinc$	−18 824.37 (4 875.32)	9 397.94 (4 812.22)	3 780.34 (4 453.42)
$y81 \cdot nearinc$	−11 863.90 (7 456.65)	−21 920.27 (6 359.75)	−14 177.93 (4 987.27)
其他控制变量	无	age，age^2	全部
观测个数	321	321	321
R-平方	0.174	0.414	0.660

我们无法从方程（13.4）和方程（13.5）得到的一个数就是 $\hat{\delta}_1$ 的标准误。$\hat{\delta}_1$ 的 t 统计量约为 -1.59，它相对于单侧对立假设来说，仅达到显著水平 5%的边缘（p 值≈0.057）。

基尔和麦克莱恩（Kiel and McClain，1995）在他们的焚化炉坐落分析中，把住房的种种特征都包括进来。这样做有两个很好的理由。首先，1981 年出售的住房种类可能和 1978 年出售的已经有了系统的差别；如果是这样，则控制住那些可能有所不同的特征就是重要的。其次，即使住房特征值平均地说在两个年份里是一样的，把它们包括进来却能大大减少误差方差，从而降低 $\hat{\delta}_1$ 的标准误。（为了讨论，可参考 6.3 节。）在第（2）列中，我们用一个二次式控制着住房的年岁（新旧程度），这将大大提高 R-平方（由于降低
417 残差方差）。现在，$y81 \cdot nearinc$ 的系数已大幅度提高，而其方差则变低。

除了在第（2）列控制的年岁变量外，第（3）列还控制了以公尺计的到达高速公路的距离（*intst*）、以公尺计的土地面积（*land*）、房间数（*rooms*）和浴间数（*baths*）。由此得到 $y81 \cdot nearinc$ 的系数估计值与没有任何控制时的系数估计值很接近，但相应的标准误却小多了：$\hat{\delta}_1$ 的 t 统计量约为

−2.84。于是在第（3）列我们得到一个比第（1）列显著得多的效应。我们选取第（3）列的估计值，因为这一列控制了最多的因素而且有最小的标准误（除了在本例中无关重要的常数项）。在第（3）列 *nearinc* 有一个小得多的系数而且是不显著的，这一事实表明了第（3）列所包含的特征基本上概括了决定房价的最重要住房特征。

为达到方法介绍的目的，在表 13.2 中我们使用了真实房价的水平值。如果使用 log(*price*)［或 log(*rprice*)］来做分析，以便得到一个近似的百分比效应，就会更有意义。这时，基本模型变为

$$\log(price) = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 \cdot nearinc + u \qquad (13.8)$$

现在，$100 \cdot \delta_1$ 是由焚化炉引起的近似的房价百分比跌落。［一如例 13.2，用 log（*price*）而不用 log（*rprice*）仅仅影响 *y81* 的系数。］使用同样的 321 个混合观测值给出

$$\begin{aligned} \log(price) = {} & \underset{(0.31)}{11.29} + \underset{(0.45)}{0.457}\, y81 - \underset{(0.055)}{0.340}\, nearinc - \underset{(0.083)}{0.063}\, y81 \cdot nearinc \\ & n = 321, R^2 = 0.409 \end{aligned} \qquad (13.9)$$

交互作用项的系数示意我们，由于新焚化炉的建造，靠近它的住房约损失 6.3% 的价值。然而，这个估计值却不是统计上异于零的。但当我们利用全部控制变量时，如表 13.2 第（3）列所示（注意 *intat*，*lana* 和 *area* 均以对数形式出现），*y81* · *nearinc* 的系数将变为 −0.132，其 *t* 统计量约为 −2.53。再次说明，控制好其他因素看来是重要的。使用对数形式，我们估计靠近焚化炉的住房贬值了大约 13.2%。

用于上例的一般性方法有着许许多多的用途，尤其是当数据来自于**自然实验**（natural experiment）或**准实验**（quasi-experiment）的时候。自然实验出现于某些外生事件——常常是政府的政策改变——改变了个人、家庭、厂商或者城市运作的环境之际。一个自然实验总有一个不受政策变化影响的对照（control）组和一个被认为受政策变化影响的处理（treatment）组，它不同于真实实验（true experiment）。在真实实验中，处理组和对照组是随机而明确地抽取的；而在自然实验中，对照组和处理组均来自某个具体的政策变化。为了控制好对照组和处理组之间的系统差异，我们需要两个年份的数 418 据，一个在政策改变以前，另一个在政策改变之后。于是我们的样本就按使用目的划分为 4 组：变化前的对照组，变化后的对照组，变化前的处理组和变化后的处理组。

把对照组称做 A，处理组称做 B，并令 *dB* 等于 1，如果预测对象属于 B；否则等于零。再令 *d2* 为第 2（政策改变后的）时期的虚拟变量，我们感兴趣的方程是

$$y = \beta_0 + \delta_0 d2 + \beta_1 dB + \delta_1 d2 \cdot dB + other\ factors \qquad (13.10)$$

式中，y 为我们感兴趣的结果变量。像例 13.3 那样，δ_1 度量着政策的效果。如果回归中没有其他因素，$\hat{\delta}_1$ 就是差异中的差分估计量：

$$\hat{\delta}_1=(\bar{y}_{2,B}-\bar{y}_{2,A})-(\bar{y}_{1,B}-y_{1,A}) \tag{13.11}$$

式中，字母上方的一横表示平均，第一个下标表示年，第二个下标表示组。当我们在方程（13.10）中增加解释变量（以便控制被抽样的总体在这两个时期之间可能有系统的变化）时，δ_1 的 OLS 估计值不再有像式（13.11）那样简单的形式，但有类似的含义。

例 13.4　工人劳工补偿法对久期的影响

迈耶、维斯卡西和德宾（Meyer, Viscusi and Durbin, MVD, 1995）研究一名受伤工人领取劳工补偿金的持续时间。在 1980 年 7 月 15 日那天，肯塔基州提高了劳工补偿金所支付的每周工资的封顶值（cap on weekly earning）。封顶值的这一提高理应对低收入工人的福利没有影响，但它将使高收入工人继续领取劳工补偿金时付出较少的代价。因此，低收入工人构成了对照组，而高收入工人构成了处理组；我们把高收入工人定义为受政策变化前封顶值约束的那些人。MVD 利用政策变化前后的两个随机样本，检验了更丰厚的劳工补偿金是否延长了人们离开工作岗位的时间（其他条件均不变）。他们首先以 log(*durat*)(表示持续时间的对数）为因变量做一个差异中的差分分析。令 *afchnge* 为政策变化后观测的虚拟变量，而 *lighearn* 为高收入者的虚拟变量。带有括号中附有标准误的估计方程为

$$\begin{aligned}\log(\widehat{durat}) = {} & 1.126+0.007\,7afchnge+0.256highearn\\ & (0.031)(0.044\,7) \qquad\qquad (0.047)\\ & +0.191afchnge\cdot highearn \\ & \quad(0.069)\\ & n=5\,626, R^2=0.021\end{aligned} \tag{13.12}$$

因此，$\hat{\delta}_1=0.191$（$t=2.77$），意味着由于收入封顶值的提高，领取劳工补偿金的平均时间延长了约 19%。*afchnge* 的系数既小且统计不显著：有如所料，收入封顶值的增加对低收入工人无延长效应。 419

尽管我们还不能对因变量的变异作出很多的解释，但为了怎样能对一种政策变化的效应得到一个较准确的估计，这却是一个不错的例子。式（13.12）中的虚拟变量仅解释了 log（*durat*）变异的 2.1%。这示意我们，显然还有许多许多影响某人领取劳工补偿金多久的因素，其中包括受伤的严重性，幸而我们用了一个很大的样本，从而得到了一个显著的 t 统计量。

MVD 曾增加各种解释变量以控制性别、婚姻状况、年龄、行业和工伤种类，得以说明在这两年里工人的情况和工伤的种类有系统的差异。但控制了这些因素，仍然看不出对 δ_1 的估计有什么影响，参看习题 13.10。

有时，两组人群由居住在美国两个邻州的人构成。例如，为了评价改变

对香烟消费的课税，可以在两个年份里分别从两个州抽取随机样本。州 A 作为对照组，没有改变香烟税，州 B 这两年间增加（或减少）了课税。结果变量将是关于香烟消费的一个度量。从而方程（13.10）可用来估计课税对香烟消费的影响。

问题 13.2 从方程（13.12）中 *highearn* 的系数及其 t 统计量，你会作什么解释？

关于自然实验方法论和更多的一些例子，可参考迈耶（Meyer，1995）的一篇有趣的综述文章。

13.3 两时期综列数据分析

现在我们转到最简单的综列数据分析：对个人、学校、厂商、城市或别的什么一个横越面，我们有了两年的数据，称之为：$t=1$ 和 $t=2$。这两年不一定要相邻的。但 $t=1$ 对应于较早的一年。例如，文件 CRIME2. RAW 含有 1982 年和 1987 年的犯罪和失业（以及其他方面）的数据，因此 $t=1$ 对应于 1982 年，而 $t=2$ 对应于 1987 年。

如果我们用 1987 年为横截面数据做一个 *crmrte* 对 *unem* 的回归而得到

$$
\begin{aligned}
\widehat{crmrte} &= 128.28 - 4.16\,unem \\
&\quad\ (20.76)\quad (3.42) \\
&\quad\ n = 46, R^2 = 0.033
\end{aligned}
$$

那么发生了什么？如果我们不经心地解释这个估计的方程，它就意味着增加失业率会降低犯罪率，这显然不是我们所预料的。*unem* 的系数在标准的显著性水平上还不是统计上显著的：至少我们还没有找到犯罪率和失业率的联带关系。

420 如本书一直所强调的，这个简单的回归方程很可能遇到遗漏变量的问题。一个可能解决的方法是试图控制更多的因素诸如年龄分布、性别分布、教育水平、执法力度等等。在多元回归分析中有许多因素可能难以控制。在第 9 章我们曾表明，怎样把以前某年（本例中就是 1982 年）的 *crmrte* 包括到分析中来，就能有助于控制不同城市在历史上有不同的犯罪率这一事实。这是利用两个年份的数据来估计因果性效应的一种方法。

利用综列数据的另一方法，是把影响因变量的观测不到的因素分为两类：一类是恒常不变的；另一类则随时间而变。令 i 表示横截面单元，t 表示时期，可将观测到的单一个解释变量的模型写成

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it},\quad t = 1,2 \tag{13.13}$$

在符号 y_{it} 中，i 表示个人、厂商、城市等等，而 t 表示时期。变量 $d2_t$ 是当

$t=1$时等于零而当 $t=2$ 时等于 1 的一个虚拟变量；它不随 i 而变，这也说明它为什么没有下标 i。因此，$t=1$ 的截距是 β_0，而 $t=2$ 的截距离是 $\beta_0+\delta_0$。一如使用独立混合横截面的情形，容许截距随时间而变，对大多数应用来说都是重要的。在犯罪一例中，在美国有一种长期趋势使得美国的每个城市的犯罪率都在改变，也许经过 5 年就看出有明显的趋势。

变量 a_i 概括了影响着 y_{it} 的全部观测不到的、在时间上恒定的因素。（a_i 没有下标 t 这一事实就告诉我们它不随时间而变。）通常把 a_i 称做**非观测效应**（unobserved effect）。在应用研究中也常常把 a_i 当做一种**固定效应**（fixed effect），这有助于我们记住 a_i 在时间上是固定的，式（13.13）中的模型被称为非观测效应模型或固定效应模型。在应用中，你也许还看到 a_i 又指**非观测差异性**（unobserved heterogeneity）（或者个人差异性、厂商差异性、城市差异性，如此等等）（注意：非观测效应模型和固定效应模型并非等同——译者注）。

误差 u_{it} 常被称做**特异性误差**（idiosyncratic error）或**时变**（time-varying）误差，因为它代表因时而变且影响着 y_{it} 的那些非观测因素。这和纯粹时间序列回归方程中的误差非常相像。

用以描述 1982 年和 1987 年城市犯罪率的一个简单的非观测效应模型是

$$crmrte_{it}=\beta_0+\delta_0 d87_t+\beta_1 unem_{it}+a_i+u_{it} \tag{13.14}$$

式中，$d87$ 为代表 1987 年的虚拟变量。由于 i 表示不同的城市，我们把 a_i 称做非观测城市效应或城市固定效应：它代表了影响城市犯罪率的、不随时间而变的全部因素。诸如城市位于美国的某一区域等等地理特征，均包含于 a_i 之中。其他的许多因素不一定毫无变化，但在 5 年期间，可视为大致上不变。后者可包括居民的某些人口特性（年龄、种族和教育），不同城市对犯罪的报导有其各自的方法，而且生活在不同城市的居民对待犯罪会有不同的态度。这些却是变化缓慢的。出于历史上的原因，不同城市可能有很不相同的犯罪率，这些现象至少部分地由非观测效应 a_i 加以描述了。

给定两年的综列数据，怎样估计我们感兴趣的参数 β_1？一种可能是基
421 本上和 13.1 节所做的那样，仅仅把两年的数据混合起来，然后用最小二乘法。这种方法有两个缺点，最重要的一点是，为了使混合的 OLS 产生 β_1 的一个一致估计量，我们就必须假定非观测效应 a_i 与 x_{it} 无相关关系。我们能容易看到这点，只要把式（13.13）写成

$$y_{it}=\beta_0+\delta_0 d2_t+\beta_1 x_{it}+v_{it},\quad t=1,\ 2 \tag{13.15}$$

式中，$v_{it}=a_i+u_{it}$ 常被称做**复合误差**（composite error）。从有关 OLS 的知识看，为了使 OLS 一致性地估计 β_1（以及其他参数），必须假定 v_{it} 与 x_{it} 不相关，这里 $t=1$ 或 2。这样说，且无论我们使用单一个横截面还是混合使用两个横截面，都是对的。因此，即使我们假定特异误差 u_{it} 与 x_{it} 不相关，如果 a_i 与 x_{it} 相关的话，混合 OLS 估计就是偏误且不一致的。由此造成的偏误有时又称做**差异性偏误**（heterogeneity bias），然而，它的确是由于遗漏了一个时间上恒常的变量而引起的。

问题 13.3 ☞

假定 a_i，u_{i1}和 u_{i2}都有零均值且两两不相关，说明：$\mathrm{Cov}(v_{i1}, v_{i2}) = \mathrm{Var}(a_i)$，从而，除非 $a_i = 0$，复合误差在时间上有正的序列相关。这对于来自混合 OLS 估计的通常 OLS 标准误来说意味着什么？

为说明所出现的情况，我们用 CRIME2·RAW 中的数据对式（13.14）做混合 OLS 估计。因有 46 个城市和每个城市各两年的数据，我们共有 92 个观测值：

$$\begin{aligned} \widehat{crmrte} &= 93.42 + 7.94d87 + 0.427unem \\ &\quad (12.74)(7.98) \qquad (1.188) \\ &\quad n = 92, R^2 = 0.012 \end{aligned} \tag{13.16}$$

（在报告所估计的方程时，通常都把下标 i 和 t 略去。）式（13.16）中 *unem* 的系数虽然是正的，其 t 统计量却非常之小。可见，使用混合两年的 OLS 和使用单一个横截面相比没有什么实质性的变化。因为使用混合的 OLS 并不解决遗漏变量的问题，所以没有什么可奇怪的。（由于前面指出的序列相关性，方程中的标准误是不正确的，但因混合 OLS 不是这里讨论的焦点，我们暂且忽略这个问题。）

在大多数应用中，收集综列数据的主要理由是为了考虑非观测效应与解释变量相关。例如，在犯罪方程中，我们有意让 a_i 中的未测出的却影响着犯罪率的因素也与失业率相关。事实上这是容易处理的：因为 a_i 在时间上是恒常的，我们可以取两个年份的数据之差。说得更准确些，对横截面第 i 个观测值，把两年的方程分别写为

$$y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2}, \quad t = 2$$

$$y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1}, \qquad t = 1$$

从第一个方程减去第二个，便得到

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

422 或

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i \tag{13.17}$$

式中，“Δ”表示从 $t = 1$ 到 $t = 2$ 的变化。非观测效应 a_i 不再出现于式（13.17）：它已被“差分掉”了，而且式（13.17）中的截距实际上是截距从 $t = 1$ 到 $t = 2$ 的变化。

方程（13.17），我们称之为**一阶差分方程**（first-differenced equation），是一个很简单的方程。它不外是单一横截面方程，仅仅每个变量取其时间上的差分而已。我们可以用第一篇所讲的方法去分析式（13.17），只要基本的假定得到满足。其中最重要的假定是 Δu_i 与 Δx_i 不相关。如果在每个时期 t，特异误差 u_{it}都与两个时期的解释变量不相关，这个假定就是对的。这是我们在第 10 章讲时间序列模型时遇到过的**严格外生性**（strict exogeneity）假

定。特别，这个假定排除了代表滞后因变量的情形。不同于第10章，我们容许与时间上恒定的不可观测量相关。当我们求出（13.17）中 β_1 的OLS估计量时，我们把所得到的估计量称做**一阶差分估计量**（first-differenced estimator）。

在犯罪一例中，假定与不相关虽然说得过去，但也可能不对。例如，假令执法力度在失业率减少的城市中要加大一些，这就造成 Δu_i 与 $\Delta unem_i$ 之间的负相关，从而导致OLS估计量的偏误。自然，这个问题可以在方程中加进更多的因素而在某种程度上加以克服，这是我们以后要讲到的。和以往情况一样，我们未能把足够多的随时间而变的因素考虑进来，这总是可能的。

另一关键性条件是，Δx_i 必须因 i 的不同而有所变化。如果解释变量对任何一次横截面观测来说都不随时间而变，或者在每一观测中都出现等量的变化，这一条件便不成立。对犯罪一例来说，因为所有城市的犯罪率都随时间而变，所以这不是什么问题。但若 i 代表个人，而 x_{it} 是一个代表性别的虚拟变量，则对一切 i 都有 $\Delta x_i = 0$；这时我们显然不能用OLS来估计式（13.17）。这点其实可以圆满地加以解释：由于容许 a_i 与 x_{it} 相关，就不要指望能把 a_i 对 y_{it} 的影响与不随时间而变的任何变量的影响分离开来。

我们需要用于通常OLS统计量的、惟一的其他假定就是式（13.17）要满足同方差性。在许多情况下这个假定是合理的。并且，如果该假定不成立的话，也知道怎样用第8章的方法去检验并更正异方差性。这样一来，OLS估计量就是无偏的，并且全部统计推断都是准确的了。

现在我们以犯罪率的例子为例估计式（13.17），得到

$$\begin{aligned} \Delta \widehat{crmrte} &= \underset{(4.70)}{15.40} + \underset{(0.88)}{2.22}\Delta unem \\ n &= 46, R^2 = 0.127 \end{aligned} \tag{13.18}$$

428 终于给出了犯罪率与失业率之间的一个正的、统计上显著的关系。由此可见，在本例中差分掉时间上恒定的效应，使得结果大不相同。式（13.18）的截距也告诉我们一些有趣的事情：即使 $\Delta unem = 0$，我们能预料犯罪率（每千人的犯罪次数）增加15.40。这反映了1982—1987年全美犯罪率的长期增长。

即使我们不从非观测效应模型（13.13）开始，利用跨时间差分仍有其直观意义。我们不去估计一个标准的横截面关系式——这会遇到遗漏变量的困扰，以致难以作出其他条件不变的结论——而是通过方程（13.17），显式地考虑解释变量在时间上的变化如何影响同期间的 y 的变化。不管怎样，把式（13.13）记在心中仍然是很有用的，它明白地表示，我们能够在保持 a_i 不变的情形下估计 x_{it} 对 y_{it} 的影响。

虽然取两年综列数据的差分是控制非观测效应的有力方法，却要付出代价。首先，综列数据比单一个横截面更难以收集，特别是关于个人的数据。我们必须进行一次调查，然后跟踪所有的个人再进行另一次跟踪调查。在第

二次调查中要确定某些人的地址常常是困难的。可以利用一个大的横截面以克服这一困难，但这又不总是可能的。而且，利用较长时间间隔的差分有时比利用逐年的变化来得好。

作为一个例子，考虑估计教育回报的问题，且使用关于个人的两年综列数据。第 i 个人的模型是

$$\log(wage_{it}) = \beta_0 + \delta_0 d2_t + \beta_1 educ_{it} + a_i + u_{it}, \quad t = 1,2$$

式中，a_i 包含非观测（即观测不到的）能力——能力大概与 $educ_{it}$ 相关。我们仍然考虑不同时期有不同的截距，用以说明总生产力增长（以及通货膨胀，如果工资指的是名义工资的话）。因为按定义天资（天生能力）是不随时间而变的，所以综列数据方法看来对教育回报的估计是合乎理想的。一次差分方程是

$$\Delta\log(wage_i) = \delta_0 + \beta_1 \Delta educ_i + \Delta u_i \tag{13.19}$$

并且可用 OLS 进行估计，问题在于，我们感兴趣的是正在工作的成年人，而大多数就业的个人在不同时期里的教育水平并没有改变。如果我们的样本中只有一小部分人有不同于零的 $\Delta educ_i$ 值，要想从式（13.19）得到 β_1 的准确估计值就是困难的，除非我们有一个相当大的样本。用一阶差分方程来估计教育回报，理论上是个好主意，但凭现有的大多数综列数据库，这又是难以付诸实施的。

424 增加一些解释变量不致引起什么困难，我们从非观测效应模型

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + a_i + u_{it} \tag{13.20}$$

开始。这个方程因为每个解释变量有三个下标，看起来比实际上要更复杂些。第一个下标指横截面观测单元，第二个下标指时期，而第三个下标不外是一个变量的标志。

例 13.5　睡眠与工作比较

我们利用来自于比德尔和哈默什（Biddle and Hamermesh，1990）的 SLP75-81.RAW 中的两年综列数据，估计睡眠与工作之间的替换关系。在问题 3.3 中，我们仅用了 1975 年的横截面。1975 年和 1981 年两年的综列数据集含有 239 人，比含有 700 人的 1975 年的横截面要小得多。以分钟计的每周睡眠总时间的一个非观测效应模型是

$$\begin{aligned} slpnap_{it} &= \beta_0 + \delta_0 d81_t + \beta_1 totwrk_{it} + \beta_2 educ_{it} + \beta_3 marr_{it} \\ &\quad + \beta_4 yngkid_{it} + \beta_5 gdhlth_{it} + a_i + u_{it}, \quad t = 1,2 \end{aligned}$$

非观测效应 a_i 可称非观测个人效应或个人固定刻画。容许 a_i 与 $totwrk_{it}$ 相关有着潜在的重要性：影响着人们睡眠多少（为 a_i 所描绘）的因素（有些是生理上的）也是那些很可能与花在工作上的时间相关的因素。有些人就是精力比较充沛，使得他们睡眠较少而工作较多。变量 *educ* 是受教育年数，*marr* 是婚姻虚拟变量，*yngkid* 是用以表示有无幼童的虚拟变量，而 *gdhlth*

是“健康良好”虚拟变量，注意，性别或种族不随时间而变，所以我们不把它们包括进来（这和横截面分析不一样）。我们的主要兴趣在于 β_1。

取两年间的差分给出可估方程

$$\begin{aligned}\Delta slpnap_t &= \delta_0 + \beta_1 \Delta totwrk_t + \beta_2 \Delta educ_t + \beta_3 \Delta marr_t \\ &\quad + \beta_4 \Delta yngkid_t + \beta_5 \Delta gdhlth_t + \Delta u_t\end{aligned}$$

假定特异误差的改变量 Δu_t 与所有解释变量的改变量都不相关，我们便能用 OLS 得到无偏估计量。这给出

$$\begin{aligned}\Delta s\hat{l}pnap = {} & 92.63 \quad 0.227\Delta totwrk - 0.024\Delta educ \\ & (45.87)(0.036) \qquad\quad (48.759) \\ & + 104.21\Delta marr + 94.67\Delta yngkid + 87.58\Delta gdhlth \\ & (92.86) \qquad\quad (87.65) \qquad\quad (76.60) \\ & n = 239, R^2 = 0.150 \end{aligned} \tag{13.21}$$

$\Delta totwrk$ 的系数表明睡眠与工作之间的替换关系：保持其他因素不变，每多工作一小时，睡眠就要减少 0.227（60）＝13.62 分钟。t 统计量(－6.31)非常显著。除截距外，再没有其他估计值是显著的了。对除 $\Delta totwrk$ 以外的全部变量的联合显著性 F 检验说，p 值＝0.49，意味着它们在任何合理的显著性水平上都是联合地不显著的，从而可把它们从方程中去掉。 425

$\Delta edac$ 的标准误相对于其估计值来说特别大，这是先前对工资方程所描述的现象。在 239 人的样本中有 183（占 76.6％）人所受的教育在这 6 年期间没有变化；有 90％的人最多只多受了一年教育。从 $\hat{\beta}_2$ 的极其大的标准误所反映的情况看，教育的变异太少，以致 $\hat{\beta}_2$ 的估计值没有任何精确度。但不管怎样，$\hat{\beta}_2$ 实际上是很小的。

综列数据还可用来估计有限的分布滞后模型。即使我们只设定两年的方程，我们也还需要收集多年的数据才能处理其中的滞后解释变量。下面是一个简单的例子。

例 13.6　犯罪清除率的分布滞后

埃伊德（Eide，1994）利用挪威警察管辖地区的综列数据估计一个犯罪率的分布滞后模型。惟一的解释变量是清除百分比（clear-up percentage，clrprc）——代表罪犯被判刑在犯罪案件中所占的百分比。犯罪率数据来自 1972 年和 1978 年两年。因为过去的清除率很可能对当前的犯罪有一种抑制作用。我们仿照埃伊德，把 clrpre 滞后 1 年和 2 年。这就导致如下的两年非观测效应模型：

$$\log(crime_{it}) = \beta_0 + \delta_0 d78_t + \beta_1 clrprc_{i,t-1} + \beta_2 clrprc_{i,t-2} + a_i + u_{it}$$

将方程取差分后再用 CRIME3.RAW 中的数据去估计它，得

$$\Delta\log(c\hat{r}ime) = 0.086 - 0.004\,0\Delta clrprc_{-1} - 0.013\,2\Delta clrprc_{-2}$$

$$(0.064)(0.004\,7) \qquad\qquad (0.005\,2)$$
$$n = 53, R^2 = 0.193, \bar{R}^2 = 0.161 \qquad (13.22)$$

第 2 个滞后是负的并且统计上显著，这意味着两年前提高清除率对本年的犯罪有抑制作用。具体地说，两年前的 *clrprc* 增加 10 个百分点，会导致本年度的估计犯罪率有 13.2% 的下跌。这表示多用警察资源及时判处罪犯可以减少未来的犯罪。

综列数据的编排

在计量经济研究中使用综列数据时，需要知道这些数据是怎样储存的。我们必须注意数据的编排，以使相同的横截面单元（个人、厂商、城市等 426 等）在不同时期能容易地连接起来。具体地说，假使我们有关于城市的两个不同年份的数据。为了大多数的目的，登记数据的最好方法是，对每个城市都安排两个记录，每年一个，每个城市的第一个记录对应于较早的一年，第二个记录对应于较晚的一年。这两个记录应该放在相邻位置（如相邻的两行）。这样，100 个城市各两年，就包含 200 个记录。头两个记录用于样本中第一个城市，下两个记录用于第二个城市，余类推。（作为一个例子，参看第 1 章的表 1.5。）这样就容易构造差分并把这些差分储存在每个城市的第二个（如第二行）记录中，以便于做混合横截面分析，并且和差分估计相比较。

伴随着本教材的两期综列数据集，大多数都是按此方式储存的（如 CRIME2.RAW，CRIME3.RAW，GPA3.RAW，LOWBRTH.RAW 和 RENTAL.RAW）。对于多于两个时期的综列数据，我们采用这种方法的直接推广。

编排两期综列数据的第二种方法是，对每个横截面单元仅安排一个记录。这就需要对每个变量做两个载入，每时期一个，SLP75-81.RAW 中的综列数据就是这样编排的。每人都有关于变量 *slpnap*75，*slpnap*81，*totwrk*75，*totwrk*81 等等的数据。构造从 1975 年到 1981 年的差分是容易的。属于这种结构的其他综列数据集还有 TRAFFIC1.RAW 和 VOTE2.RAW。把数据都安排在一个记录里的缺点是无法按照（指电脑）两期的原始数据进行混合 OLS 分析。而且这种编排方法不适应于多于两期的综列数据集，后者是我们将要在 13.5 节中考虑的问题。

13.4 用两期综列数据做政策分析

综列数据对于政策分析非常有用，特别是计划（项目）的评估。最简单

的计划评估的格局是，先在第一个时期里抽取一个由个人、厂商或城市以至其他单元组成的样本。然后让这些单元的一部分参与以后的一个时期举办的某个计划（项目），那些不参加计划的单元则作为对照组。这和以前讨论过的自然实验相像，但有一点重要的差别：在每一个时期里我们都看到同样的横截面单元。

作为一个例子，假定我们想评估一下，密歇根州在职培训计划对制造业厂商的劳工生产率的影响（并参见问题9.8）。令 $scrap_{it}$ 表示第 t 年第 i 个厂商的报废率（每100件中由于有缺陷而必须报废的件数），$grant_{it}$ 是二值标示变量；它等于1，如果在第 t 年厂商 i 领取一项在职培训津贴。对1987年和1988两年，模型是

$$scrap_{it} = \beta_0 + \delta_0 y88_t + \beta_1 grant_{it} + a_i + u_{it}, \quad t = 1,2 \tag{13.23}$$

式中，$y88_t$ 是对1988年的虚拟变量，而 a_i 是非观测厂商效应或厂商固定效应。非观测效应包括平均雇工能力、资本、管理技能，等等；这些因素在这两年期间基本上保持不变。我们关注 a_i 可能系统地与厂商是否领取这项津贴有关。例如，计划执行人对工人技能较低的厂商给予优先考虑。或者，恰恰相反，执行人为了显示在职培训计划的有效性，也许把津贴奖给有较高劳
427 工生产率的厂商（雇主），事实上，在本计划中，津贴的发放是按先来先得的原则。但是，一个厂商是否及早申请津贴也许和他的工人生产率有关。不管怎样，用单一个横截面或混合两个横截面的分析都将产生偏误的和非一致的估计量。

取差分以消去 a_i，给出

$$\Delta scrap_i = \delta_0 + \beta_1 \Delta grant_i + \Delta u_i \tag{13.24}$$

因此，只须求报废率的改变量对津贴标示量的改变量的回归即可。因1987年尚无厂商领取津贴，即对所有的 i，$grant_{i1} = 0$，故 $\Delta grant_i = grant_{i2} - grant_{i1} = grant_{i2}$，这无非表示一个厂商在1988年是否领取了津贴。然而，一般地说，由于必须消去非观测效应模型中的 a_i，取所有变量（包括虚拟变量）的差分是重要的。

利用JTRAIN.RAW中的数据估计一阶差分方程，得到

$$\begin{aligned}\Delta \hat{scrap} &= -0.564 - 0.739\Delta grant\\ &\quad\;(0.405)\quad (0.683)\\ &\quad n = 54, R^2 = 0.022\end{aligned}$$

因此，我们估计，得到一项在职培训津贴平均降低报废率 -0.739。但这个估计值并非统计上异于零。

使用log(*scrap*)估计百分率效应，得到较强的结果：

$$\begin{aligned}\Delta \log(\hat{scrap}) &= -0.057 - 0.317\Delta grant\\ &\quad\;(0.097)\quad (0.164)\\ &\quad n = 54, R^2 = 0.067\end{aligned}$$

即得到一项在职培训津贴估计能把报废率降低约 27.2%。[因为$\exp(-0.317)-1\approx-0.272$]。$t$ 统计量约为 -1.93，在显著界线的边缘上。作为对比，$\log(scrap)$对 $y88$ 和 $grant$ 的回归的混合 OLS 估计给出 $\hat{\beta}_1=0.057$（标准误 $=0.431$）。这里，我们没有发现报废率与在职培训津贴之间的任何显著关系。由于这一结果与一阶差分估计相差甚远，这就示意我们能力低的工人更有可能领取到一份津贴。

对计划评估模型作更一般的研究，似乎是有用的。令 y_{it} 为结果变量，并令 $prog_{it}$ 为计划参与虚拟变量。最简单的非观测效应模型是

$$y_{it}=\beta_0+\delta_0 d2_t+\beta_1 prog_{it}+a_i+u_{it} \tag{13.25}$$

如果计划参与仅发生在第二个时期，那么在差分方程中 β_1 的 OLS 估计量就有一个非常简单的表示式：

$$\hat{\beta}_1=\overline{\Delta y}_{treat}-\overline{\Delta y}_{control} \tag{13.26}$$

428 也就是，我们计算处理组和对照组在这两个时期的平均变化，然后取两者的差就是 $\hat{\beta}_1$，这是对两个混合横截面而言，方程（13.11）所描述的差异中的差分估计量的综列数据翻版。有了综列数据，就有一个潜在的重大优势。我们可以对于同样的横截面单元取 y 在不同时期的差分，由此得以控制个人、厂商或城市特有的效应，如同式（13.25）中的模型所表明的那样。

如果计划参与发生在两个时期，$\hat{\beta}_1$ 就不能写成式（13.26）那样，但对它的解释是一样的：它代表 y 均值由于计划参与而起的变化。

控制随时间而变化的因素，不会改变任何重要的东西。我们只须取这些变量的差分，连同 $\Delta prog$ 一起把它们包括进来，这样就可以控制有可能与计划目的相关的随时间而变的变量。

同样的方法适用于分析任何因城市或州而异的政策效应。下面是一个简单的例子。

例 13.7　酒后驾驶法对交通伤亡事故的影响

为了制止酒后驾驶，美国许多州采取了不同的政策。这里我们将研究两类法规：开瓶法——规定乘客携带打开瓶的酒精饮料为非法——以及行政权力法——法院可在司机因酒后驾驶被逮捕后而未判刑前吊销其执照。一种可能做的分析是，利用州的单一个横截面将行车死亡事故变量对标示着每种法规是否出现的虚拟变量做回归。但这种分析方法不一定能奏效，因为州是按照立法程序来决定它们是否需要这些法规的。所以，法规的出现很可能与近年来酒后驾车的平均死亡事故有关。一种较有说服力的分析是使用某些州采取新法规的那段时间的综列数据（而另一些州也许在这一时期里废除了现行的法规）。文件 TRAFFIC1.RAW 含有全部 50 个州和哥伦比亚特区的 1985 年和 1990 年两年数据。因变量是每行驶 1 万公里的交通事故致死人数（*dthrte*）。在 1985 年有 19 个州采用开瓶法，而 1990 年有 22 个州采用这种法规。在 1985 年有 21 个州采用权力法，到了 1990 年采用权力法的州增加

到 29 个。

取一阶差分，后利用 OLS 得到

$$\Delta dthrte = -0.497 - \underset{(0.052)(0.206)}{0.420}\Delta open - \underset{(0.117)}{0.151}\Delta admn \qquad (13.27)$$

$$n = 51, R^2 = 0.119$$

这些估计值表示，采取开瓶法，把交通事故死亡率降低了 0.42；给定 1985 年的平均死亡率为 2.7，而其标准差为 0.6，这是一个不容忽略的下降。这个估计值相对于一个双侧对立假设来说是在 5% 水平上统计显著的。权力法则有较小的效应，其 t 统计量仅为 −1.29；但有我们所预期的符号，此方程的截距表明，无论法规有无变化，所有的州在这 5 年期间的交通事故死亡率均已大大降低。那些采取开瓶法的州的死亡率平均地说，则有更进一步的降低。429

问题 13.4

在例 13.7 中，对华盛顿州有 $\Delta admn = -1$，这意味着什么？试加解释。

其他法规如座位绑带法、摩托驾车带盔法、最高速限法也可能影响交通事故死亡。此外，还可以考虑对年龄和性别分布的控制，以及各州的诸如“母亲反对酒后驾驶”一类组织的影响。

13.5 多于两期的差分法

差分方法还可用于多期数据。为说明起见，假定我们有 N 个个人且每人有 $T=3$ 个时期的数据，一般的固定效应模型是，对于 $t=1$，2 和 3，有

$$y_{it} = \delta_1 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it} \qquad (13.28)$$

（因此观测的总次数是 $3N$。）注意，现在除截距外，包含了两个时期的虚拟变量。对每一时期分别考虑一个不同的截距这个做法是可取的，尤其是在为数不多的时期的情况下。如同平常那样，基期是 $t=1$。第二时期的截距是 $\delta_1+\delta_2$，如此类推，我们主要感兴趣的是 β_1，β_2，…，β_k。如果非观测效应 a_i 与任一解释变量相关，则对这 3 年数据使用混合 OLS 将导致偏误且非一致估计值。

关键的一个假定是，特异误差与每一时期的解释变量都不相关：

$$\mathrm{Cov}(x_{itj}, u_{is}) = 0 \qquad (13.29)$$

也就是，当我们把非观测效应 a_i 去掉后，诸解释变量都是严格外生的。（用零条件期望值陈述的严格外生性假定，见于章末附录。）假定式（13.29）排

除了将来的解释变量会对特异误差的当前变化有所回应的情形；但当 x_{itj} 是滞后因变量时，这种情形则必定发生。如果我们遗漏了一个重要的随时间而改变的变量，那么一般地说式（13.29）是不成立的。一个或多个解释变量有测量误差时，正如我们在第9章中看到的那样，也会使式（13.29）成为谬误。对于这种情形，如何补救，将在第15、16章中讨论。

如果 a_i 与 x_{itj} 相关，则 x_{itj} 将与复合误差 $v_{it}=a_i+u_{it}$ 相关。我们可以取相邻期的差分把 a_i 去掉。对于 $T=3$ 的情形，我们从第二期减去第一期，并从第三期减去第二期，即

$$\Delta y_{it}=\delta_2\Delta d2_t+\delta_3\Delta d3_t+\beta_1\Delta x_{it1}+\cdots+\beta_k\Delta x_{itk}+\Delta u_{it} \qquad (13.30)$$

430 式中 $t=2$ 和3。当 $t=1$ 时，因没有可从 $t=1$ 的方程中减去的东西，所以就没有 $t=1$ 的差分方程。现在，式（13.30）就代表样本中每一个人各两个时期的方程。如果该方程满足经典线性模型假定，则混合OLS将给出无偏估计量，并且 t 和 F 统计量可用于假设检验。还可诉之于渐近结果。为使OLS是一致的，有一个重要的要求，就是对所有的 j 和 $t=2$ 和3，Δu_{it} 都与 Δx_{itj} 不相关。这是对两时期情形的自然延伸。

注意式（13.30）是怎样地含有年虚拟变量 $d2_t$ 和 $d3_t$ 的差分的。当 $t=2$ 时，$\Delta d2_t=1$ 而 $\Delta d3_t=0$；当 $t=3$ 时，$\Delta d2_t=-1$ 而 $\Delta d3_t=1$。因此，式（13.30）不含有截距。对某些目的如计算 R-平方来说，这是不方便的。除非原始模型（13.28）中的时间截距是我们直接感兴趣的参数——这种情形是少有的——我们不如估计一个含截距的单一个时期（常常是第三个时期）虚拟变量的一阶差分方程。换言之，该方程变为

$$\Delta y_{it}=\alpha_0+\alpha_3 d3_t+\beta_1\Delta x_{it1}+\cdots+\beta_k\Delta x_{itk}+\Delta u_{it},\quad t=2,3$$

不管哪一种形式，诸 β_j 的估计值都是一样的。

对多于3个时期的情形，方法也是类似的。如果对 N 个横截面单元的每一个都有同样的 T 期数据，我们就说数据集是**平衡综列数据**（balanced panel）：我们对所有的个人、厂商、城市等等都有同样的时期。当 T 相对于 N 而言较小时，我们应该给每时期一个虚拟变量，以说明建模时未考虑到的长期变化。因此，经过一阶差分，方程就像是

$$\begin{aligned}\Delta y_{it}=\alpha_0&+\alpha_3 d3_t+\alpha_4 d4_t+\cdots+\alpha_T dT_t+\beta_1\Delta x_{it1}+\cdots\\&+\beta_k\Delta x_{itk}+\Delta u_{it},\quad t=2,3,\cdots,T\end{aligned} \qquad (13.31)$$

在这个一阶差分方程中，每个单元 i 都有 $T-1$ 个时期的数据，总共有 $N(T-1)$ 个观测值。

假使观测值已适当编排，并且仔细地进行了差分，用混合OLS估计式（13.31）是简单的。为使一阶差分方便，数据文件应含有 NT 个记录。头 T 个记录用于第一个横截面观测，并按年代（日历）编排；其次的 T 个记录用于第二个横截面观测，也按年代编排；余类推。然后计算差分，将 $t-1$ 期到 t 期的变化储存在 t 期的记录上。因此，$t=1$ 期的差分对所有的 N 个横截面观测来说都应是缺落的。不做到这步，你就会在你的回归分析中冒着

使用假观测值的风险。比方说，如果将第 $i-1$ 个人的最后一个观测值从第 i 个人的第一个观测值中减去，就会造成一个虚假观测。如果你用差分数据做回归，而得到的报告是 NT 或 $NT-1$ 次观测，那么你是忘掉了设置 $t=1$ 的观测失落。

431 当我们使用多于两期的数据时，必须假定 Δu_{it} 是时序不相关的，才能使通常的标准误和检验统计量有效。这个假定有时是合理的，但它并不因为我们假定了原始的特异误差没有时序相关就自然成立（我们在第 14 章中将要用到的一个假定）。事实上，如果假定 u_{it} 时序不相关且有恒定方差，则可以证明 Δu_{it} 与 $\Delta u_{i,t-1}$ 之间的相关为 -0.5。如果 u_{it} 遵循一个稳定的 AR（1）模型，则将是时序相关的。只有当 u_{it} 遵循一个随机游走时，Δu_{it} 才会是时序不相关的。

检验一阶差分方程中的时序相关性是容易的。令 $r_{it}=\Delta u_{it}$ 表示原始误差的一阶差分。如果 r_{it} 遵循 AR（1）模型 $r_{it}=\rho r_{i,t-1}+e_{it}$，我们就能容易地检验 H_0：$\rho=0$。首先，通过混合 OLS 估计式（13.31）并求出残差 $\hat{r}_{it}$，然后用 $\hat{r}_{i,t-1}$ 作为一个添加的解释变量再做一次回归。$\hat{r}_{i,t-1}$ 的系数就是 ρ 的估计值，这样我们就能用平常的 $\hat{r}_{i,t-1}$ 的 t 统计量去检验 H_0：$\rho=0$。由于在使用滞后 OLS 残差，我们再次失去一个时期。例如，我们原有 $T=3$，差分方程就只有 $T=2$。时序相关性的检验不外是对一阶差分做一个（使用第三个时期的）含有滞后 OLS 残差的横截面回归。这和 12.2 节纯时间序列模型所讲的检验相类似。以后我们将给出一个例子。

通过拟差分（quasi-differencing）方程（13.31）我们能对 AR（1）时序相关的出现作出校正。[还可对式（13.31）中的第一个时期做 Prais-Winsten 变换]。不幸的是，对时间序列回归作出 AR（1）校正的软件包不能适用。标准的 Cochrane-Orcutt 或 Prais-Winsten 方法会把观测值看成对 i 和 t 的一个 AR（1）过程；但因我们假定了观测值对 i 而言是独立过程，这个方法就失去了意义。当 N 较大（而且 N 应明显地比 T 大）时，可以对 OLS 标准误进行校正计算，以适应于任意形式的时序相关（和异方差性）。关于这些问题的详细论述超出了本书的范围［参看 Wooldridge（1999，第 10 章）］，但在一些软件包中这些计算是容易的。

如果误差中没有时序相关，则通常处理异方差性的方法是适用的。既可使用第 8 章讲的 Breusch-Pagan 和 White 异方差检验，也可以计算稳健标准误。

对多于两年的综列数据取差分，如下面的例子所示，对于政策分析而言是非常有用的。

问题 13.5 ☞

Δu_{it} 中的时序相关会造成一阶差分估计量的偏误和非一致性吗？为什么时序相关是一个问题？

例 13.8　企业（特）区对失业贴补（申请）的影响

帕普克（Papke，1994）研究了印第安纳州企业划区（EZ）计划对失业

贴补申请数的影响。她分析了 1980—1988 年印第安纳州的 22 个城市。1984 年有 6 个企业区被划定，1985 年又划定了 4 个。在她的样本中有 12 个城市在这个期间是没有得到任何企业区的划定，从而这些城市可作为对照区来利用。

一个简单的政策评价模型是

$$\log(uclms_{it}) = \theta_t + \beta_1 ez_{it} + a_i + u_{it}$$

432 式中，$uclms_{it}$ 为第 t 年第 i 城市的失业贴补申请数。参数 θ_t 无非代表每个时期的一个不同的截距。一般地说，在这个时期里全州的失业贴补申请数都在下降，而且这应在不同的年截距上得到反映。二值变量 ez_{it} 等于 1，如果城市 i 在时期 t 被划为企业区；我们感兴趣的是 β_1。因为企业区的划定不是随机的——企业区通常是经济萧条的地域——所以很可能 ez_{it} 与 a_i 有正的相关。（较高的 a_i 意味着较高的失业贴补申请数，后者又导致较高的被划为 EZ 的机会。）于是，为了消掉 a_i，我们应取方程的差分：

$$\Delta\log(uclms_{it}) = \alpha_0 + \alpha_1 d82_t + \cdots + \alpha_7 d88_t + \beta_1 \Delta ez_{it} + \Delta u_{it} \quad (13.32)$$

此方程的因变量，即 log（$uclms_{it}$）的改变量，是从 $t-1$ 年到 t 年失业贴补申请数的近似年增长率。我们可以用 EZUNEM. RAM 中的数据对 1981 年到 1988 年估计这个方程；总样本大小是(22)(8) = 176。β_1 的估计值是 $\hat{\beta}_1 = -0.182$（标准误 = 0.078）。由此看来，每出现一个 EZ 将产生大约 16.6%［exp（−0.182）−1≈ 0.166］的失业贴补申请数的下降。这是一种经济上较大且统计上显著的效应。

方程中没有异方差性的迹象：Breusch-Pagan F 检验给出 $F = 0.85$，p 值 = 0.557。然而，当我们把滞后 OLS 残差加进差分方程（同时失去 1981 年）时，我们得到 $\hat{\rho} = -0.197$（$t =$ 2.44），从而看到有迹象表明在一阶误差中存在有最低的负时序相关。不同于时序正相关的情形，当误差是负相关时，平常的 OLS 标准误并不总是大大低估了正确的标准误（见 12.1 节）。因此，企业划区虚拟变量的显著性也许不会受到影响。

例 13.9　北卡罗来纳州的地县犯罪率

康韦尔和特朗布尔（Cornwell and Trumbull，1994）利用 1981—1987 年北卡罗来纳州的 90 个县的数据估计了一个非观测效应的犯罪模型；数据见于 CRIME4.RAW。这里我们是估计他们的模型的一个简化版本。我们取方程在时间上的差分以消去非观测效应 a_i。（康韦尔和特朗布尔用的是一个不同的变换，我们将在第 14 章中加以讨论。）a_i 可能包含地理区位、对犯罪态度、历史记录和报告惯例种种因素。犯罪率指每人犯罪次数，*prbarr* 是估计的逮捕概率，*prbconv* 是估计的判罪概率（假定已逮捕），*prbpris* 是坐牢时间的概率（假定已判罪），*avgsen* 是平均服刑期长，而 *polpc* 是人均警察数。按照标准的刑事计量研究，我们取所有变量的对数以估计各种弹性。我们还引进全套年虚拟变量以控制犯罪率在该州的变化趋势，我们对 1982—

433

1987 年估计了这个差分方程。圆括号中的量是平常的 OLS 标准误，方括号中的是兼对时序相关和异方差性稳健的标准误。

$$
\begin{array}{rl}
\Delta\log(crmrte) = & 0.008 - 0.100d83 - 0.048d84 - 0.005d85 \\
& (0.017)(0.024)\quad(0.024)\qquad(0.023) \\
& [0.014][0.022]\quad[0.020]\qquad[0.025] \\
& +\,0.028d86 + 0.041d87 - 0.327\Delta\log(prbarr) \\
& \quad(0.024)\qquad(0.024)\qquad(0.030) \\
& \quad[0.021]\qquad[0.024]\qquad[0.056] \\
& -\,0.238\Delta\log(prbconv) - 0.165\Delta\log(prbpris) \qquad (13.33) \\
& \quad(0.018)\qquad\qquad(0.026) \\
& \quad[0.039]\qquad\qquad[0.045] \\
& -\,0.022\Delta\log(avgsen) + 0.398\Delta\log(polpc) \\
& \quad(0.022)\qquad\qquad(0.027) \\
& \quad[0.025]\qquad\qquad[0.101] \\
& n = 540,\ R^2 = 0.433,\ \bar{R}^2 = 0.422
\end{array}
$$

三个概率变量——逮捕、判罪和坐牢时间的概率——都有预期的符号且都是统计上显著的。例如，逮捕概率每增加 1%，预料会降低犯罪率的 0.93%。平均判刑变量有一定的抑制作用，但统计上不显著。

人均警察变量的系数有些奇怪，而且象征了大多数意在解释犯罪率的研究。如果漫不经心地解释，便是说人均警察每增加 1%，犯罪率将提高约 0.4%（常用的 t 统计量非常之大，差不多是 15。）很难相信多些警察会造成更多的罪犯。到底发生了什么？至少有两种可能性。第一，犯罪率变量的计算来自于犯罪报告。也许当警察更多的时候，有更多的犯罪报告。其次，由于其他原因，方程中的警察变量是内生的；地方政府可能在它们预料犯罪率要上升时加强了警察力量。似此情形，式（13.33）就不能按因果方式加以解释了。在第 15 和 16 两章中，我们将讨论能解释这种附加内生形式的模型与估计方法。

8.3 节所讲的对异方差性的 White 检验给出 $F = 75.48$ 和 p 值 $= 0.000$，可见有很强的异方差现象。（严格地说，如果还出现有时序相关，这个检验便是无效的，但它有很强的示意性。）对 AR（1）时序相关的检验给出 $\hat{\rho} = -0.233$，$t = -4.77$，故存在有负的时序相关。方括号中的标准误是为防御时序相关和异方差性而校正过的。［我们不对此作详细讨论；其计算类似于 12.5 节所述，而且许多计量经济软件已都做此计算。更多的讨论，参见于 Wooldridge（1999，第 10 章）。］没有任何变量失掉统计显著性，但显著的抑制变量的 t 统计量却明显变小了。例如，判罪概率的 t 统计量用平常的 OLS 标准误计算是 -13.22，而用全面稳健的标准误计算则是 -6.10。等价地说，用稳健标准误来构造的置信区间（自然而然）要比用平常 OLS 标准误构造的宽得多。

▶ 小结

434 我们学习了分析独立混合横截面和综列数据集的方法。当我们在不同时期（通常是不同年份）抽取不同的随机样本时，就有了独立横截面。主要的估计方法是利用混合数据的OLS，并且有通常的推断程序可用，包括对异方差性的校正。（因为在不同时间的样本是独立抽取的，所以没有时序相关的问题。）由于时间序列的变化，我们常常考虑不同时间有不同的截距。还可以通过时间虚拟变量与某些主要变量的交互作用看这些主要变量是怎样在时间上变化的。这对于利用自然实验作政策评价研究来说尤其重要。

在应用研究尤其是政策分析中，越来越多地用到综列数据集。当我们在不同时期对相同的横截面单元进行跟踪时，就有了综列数据集。当我们认为人们、厂商、城市等等不随时间而变的非观测特点也许与我们模型中的解释变量相关，而我们又想对这些特点加以控制，这时综列数据是最为有用的。清除非观测效应的一个方法是，在相邻的时期取数据的差分。这样就能把标准的OLS分析用于这些差分上。于是，使用两期数据的结果就是对数据差分做一个横截面回归。在同方差假定下，通常的推断程序是渐近成立的；在正态性假定下还可以作出精确的推断。

对多于两期的情形，我们可以把混合OLS用于差分数据上；由于取了差分，我们失去了第一个时期。除同方差性外，我们还必须假定差分误差无时序相关，以便应用通常的 t 和 F 统计量。（章末附录开列了一个仔细的有关假定的清单。）自然，任何不随时间而变的变量都会在分析中消失。

关键术语

平衡综列（数据）
复合误差
差异中的差分估计量
一阶差分方程
一阶差分估计量
固定效应
固定效应模型
差异性偏误
特异误差
独立混合横截面
纵横数据
自然实验
综列数据
准实验
严格外生性
非观测效应
非观测效应模型
非观测差异性
年虚拟变量

习 题

13.1 假定在例 13.1 中除教育以外的所有因素的平均值都不随时间而变，并且教育的平均水平在 1972 年的样本中是 12.2，而在 1984 年的样本中是 13.3。利用表 13.1 中的估计值，估计 1972 年和 1984 年之间的平均生育率变化（不要忘记解释截距变化和教育的平均变化）。

435 13.2 利用 KIELMC.RAW 中的数据，对 1978 年和 1981 年估计了如下方程：

$$\begin{aligned}\widehat{\log(price)} &= \underset{(0.26)}{11.49} - \underset{(0.058)}{0.547}\, nearinc + \underset{(0.080)}{0.394}\, y81 \cdot nearinc \\ n &= 321,\ R^2 = 0.220\end{aligned}$$

和

$$\begin{aligned}\widehat{\log(price)} &= \underset{(0.27)}{11.18} + \underset{(0.044)}{0.563}\, y81 - \underset{(0.067)}{0.403}\, y81 \cdot nearinc \\ n &= 321,\ R^2 = 0.337\end{aligned}$$

试比较对交互作用项 $y81 \cdot nearinc$ 的估计值和方程（13.9）中的相应估计值。为什么这些估计值差别这样大？

13.3 当我们拥有两年的独立横截面（而不是综列数据）时，为什么不能使用一阶差分方法？

13.4 如果我们认为式（13.14）中的 β_1 是正的，并且 Δu_i 和 $\Delta unem_i$ 是负相关的，那么，在一阶差分方程中 β_1 的 OLS 估计量会有什么偏误？[提示：复习表 13.2]。

13.5 假使我们想估计若干个变量对年储蓄的影响，并且我们拥有 1990 年元月 31 日和 1992 年元月 31 日所收集的关于个人的综列数据。如果我们引进一个 1992 年虚拟变量并做一阶差分，那么我们还能在原模型中包含有年龄变量吗？试加解释。

13.6 在 1985 年，佛罗里达州和佐治亚州都没有法律禁止在行车的乘客厢中有开瓶的酒精饮料。到 1990 年佛罗里达州通过了这一法律，但乔治亚州则尚未通过。

(i) 假定你在 1985 年和 1990 年都能在两个州收集到驾驶年龄总体的各一个随机样本。令 $arrest$ 为二值变量；如果某人在该年因酒后驾驶而被捕，它就等于 1。在不对任何其他因素进行控制的情况下，写出可用以检验开瓶法是否减少了因酒后驱车被捕概率的线性概率模型。在你的模型中，哪个系数度量了此法律的效应？

(ii) 你为什么想在此模型中对其他因素进行控制？你想控制的会是哪些因素？

计算机习题

13.7　本题使用 FERTIL1.RAW 中的数据。

(i) 对例题 13.1 所估计的方程，检验 16 岁时的生活环境是否对生育率有影响。(以大城市为基组) 报告 F 统计量的值及其 p 值。

(ii) 检验 16 岁时所在地 (县) 区 (以南方为基组) 是否对生育率有影响。

(iii) 令 u 为总体方程中的误差项。假使你认为 u 的方差随时间而变
436 (但不随 *educ*, *age* 等等而变)，那么刻画这一特点的一个模型是

$$u^2 = \gamma_0 + \gamma_1 y74 + \gamma_2 y76 + \cdots + \gamma_6 y84 + v$$

利用这个模型去检验 u 的异方差性。[提示：你的 F 检验应有 6 和1 122个自由度。]

(iv) 在表 13.1 所估计方程中增加交互作用项 $y74 \cdot educ$, $y76 \cdot educ$, …, $y84 \cdot educ$。解释这些项代表了什么？它们是否联合地显著？

13.8　本题使用 CPS78-85.RAW 中的数据。

(i) 你怎样解释方程 (13.2) 中 $y85$ 的系数？对它有没有一种令人感兴趣的解释？(这里你要小心，你必须说明交互作用项 $y85 \cdot educ$ 和 $y85 \cdot female$。)

(ii) 保持其他因素不变，你估计一个受了 12 年教育的男子的名义工资增加了多少个百分点，试提出一种回归以得到这个估计的一个置信区间。[提示：为了得到这个置信区间，要用$y85 \cdot (educ - 12)$代替 $y85 \cdot educ$；参考例 6.3]。

(iii) 令所有的工资均以 1978 年美元计算，重新估计方程 (13.2)。具体地说，定义 1978 年的实际工资为 $rwage = wage$，而 1985 年的实际工资为 $rwage = wage/1.65$。现在估计式 (13.2) 时用 $\log(rwage)$代替$\log(wage)$。哪些系数将不同于方程 (13.2) 中的系数？

(iv) 解释为什么你在 (iii) 中的回归给出的 R-平方不同于方程 (13.2) 所给出的。[提示：两个回归的残差，从而残差的平方和是相同的。]

(v) 试描述 1978—1985 年参加工会 (union participation) 的作用起了什么变化？

(vi) 从方程 (13.2) 开始，检验会员工资差别是否随时间而变。[提示：应使用简单的 t 检验。]

(vii) 你在 (v) 和 (vi) 两部分中的发现是否相互矛盾？试解释。

13.9　本题使用 KIELMC.RAW 中的数据。

(i) 变量 *dist* 是以英尺计算的从每个住家到焚化炉位置的距离。考虑模型

$$\log(price) = \beta_0 + \delta_0 y81 + \beta_1 \log(dist) + \delta_1 y81 \cdot \log(dist) + u$$

如果建造焚化炉会减少其附近的住房价值，那么 δ_1 的符号将是什么？如果 $\beta_1 > 0$，这个符号意味着什么？

(ii) 估计 (i) 中的模型并按通常的方式报告你的结果。

（iii）在方程中增加 *age*，age^2，*rooms*，*baths*，log（*inst*），log（*land*）和 log（*area*）。现在，你对焚化炉对住房价值的影响会作出什么结论？

13.10　本题使用 INJURY.RAW 中的数据。

437 （i）使用肯塔基州的数据，增加 *male*，*married* 和一全组行业和工伤类型虚拟变量作为解释变量，重新估计方程（13.12）。在控制了这些其他因素之后，*afchnge*·*highearn* 的估计值有怎样的变化？这个估计值仍然是统计上显著的吗？

（ii）你对（i）的小 *R*-平方有什么可说的？这是否意味着这个方程是无用的呢？

（iii）用密歇根州的数据估计方程（13.12）。比较密歇根州和肯塔基州的关于交互作用项的两个估计值。密歇根州的估计值在统计上显著吗？你从这里能引伸出什么呢？

13.11　本题使用 RENTAL.RAW 中的数据。1980 年和 1990 年的数据包括各大学城的租金和其他变量。我们的意图是，看看更多学生的出现会不会影响房租。非观测效应的模型是

$$\begin{aligned}\log(rent_{it}) = \beta_0 + \delta_0 y90_t + \beta_1\log(pop_{it}) + \beta_2\log(avginc_{it})\\ + \beta_3 pctstu_{it} + a_i + u_{it}\end{aligned}$$

式中，*pop* 为城市人口；*avginc* 为平均收入；*pctstu* 为作为城市人口百分比的学生人口（按学年计算）。

（i）用混合 OLS 估计方程并按标准方式报告结果。你从 1990 年虚拟变量的估计值引出什么结论？你得到怎样的 $\hat{\beta}_{pctstu}$？

（ii）你在（i）中报告的标准误是否真实？试加解释。

（iii）现在，求方程的差分并做 OLS 估计，把你对 β_{pctstu} 的估计值和（ii）得到的相比较。学生人口的相对大小看来是否对房租有影响？

（iv）对（iii）中的一阶差分方程求对异方差性稳健的标准误。这是否改变了你的结论？

13.12　本题使用 CRIME3.RAW 中的数据。

（i）在例 13.6 的模型中检验假设 H_0：$\beta_1 = \beta_2$。［提示：定义 $\theta_1 = \beta_1 - \beta_2$ 并用 θ_1 和 β_2 来表达 β_1。以此代入方程并加以整理，对 θ_1 做 *t* 检验。］

（ii）如果 $\beta_1 = \beta_2$，差分方程就可写为

$$\Delta\log(crime_t) = \delta_0 + \delta_1\Delta avgclr_t + \Delta u_t$$

式中，$\delta_1 = 2\beta_1$，且 $avgclr_t =$（$clrprc_{t-1} + clrprc_{t-2}$）/2 是前面两年的平均清除百分比。

（iii）估计（ii）中的方程，将校正的 *R*-平方和式（13.22）中的相比较。你最后会选用哪一个模型？

13.13　本题使用 GPA3.RAW 数据。数据集来自某大学秋季和春季两个学期的 366 名学生运动员［类似的一个分析见于 Maloney 和 McCormick (1993)，但现在我们利用一个真正的综列数据集］。因为你有了每个学生的两学期数据，用一个非观测效应模型是适宜的。主要关注的问题是：运动员

们是否在他们的运动项目合时令的那个学期里功课表现较差呢?

438 (i) 用混合的 OLS 估计一个以学期总成绩 GPA（*trmgpa*）为因变量的模型。解释变量是 *spring*, *sat*, *hsperc*, *female*, *black*, *white*, *frstsem*, *tothrs*, *crsgpa* 和 *season*。试解释 *season* 的系数。它是统计上显著的吗?

(ii) 在仅参与秋季运动项目的运动员中，大多数是足球运动员。假定足球运动员的能力水平和其他运动员的能力水平有系统上的差别。如果 *SAT* 分数和中学成绩百分位数不能很好地反映一个人的能力水平，那么混合 OLS 估计量将是有偏误的。试解释为什么?

(iii) 现在，取两个学期数据的差分，问有哪些变量将随之消失? 检验是否有时令效应。

(iv) 你能想像出一个或多个有潜在重要性的、不随时间而变的变量，而在此分析中被我们忽略了的?

13.14 VOTE2.RAW 含有 1988 年和 1990 年众议院选举的综列数据。只有在 1988 年获胜并在 1990 年再次参加竞选的人才会出现在样本中；这些人是在位的竞选人。用候选双方的费用来解释在位者得票的份额的一个非观测效应模型是

$$
\begin{aligned}
vote_{it} = {} & \beta_0 + \delta_0 d90_t + \beta_1 \log(inexp_{it}) + \beta_2 \log(chexp_{it}) \\
& + \beta_3 incshr_{it} + a_i + u_{it}
\end{aligned}
$$

式中，$incshr_{it}$ 为在总竞选支出中在位者所占份额（用百分数表示）。非观测效应，包括在位者的特征——诸如品质——以及地区方面的一切，这些都不随时间而变。在位者的性别、党派都在时间上恒定，因此都属于 a_i。我们关注的是竞选费用对选举结果的影响。

(i) 取给定方程在两个年份里的差分并用 OLS 估计差分方程。问哪些变量相对于一个双侧假设是在 5% 水平上个别地显著的?

(ii) 在 (i) 的方程中，检验 $\Delta\log$（*inexp*）和 $\Delta\log$（*chexp*）的联合显著性。报告其 p 值。

(iii) 用 $\Delta incshr$ 作为惟一的自变量，重新估计 (i) 中的方程。解释 $\Delta incshr$ 的系数。例如，如果在位者的支出份额增加 10 个百分点，你预料这会怎样影响在位者的得票份额?

(iv) 再做一遍 (iii)，但现在仅限于使用双方都有重复挑战者的情形。[这样一来，我们还可以控制挑战者的属于 a_i 的那些特征。莱维特（Levitt, 1955）做过一项更为广泛分析。]

13.15 本题使用 CRIME4.RAW 中的数据。

(i) 在数据集中增加每一工资变量的对数，然后用一阶差分估计模型。问这些变量的引进怎样地影响了例 13.9 中的那些公平判刑变量的系数?

(ii) (ii) 中的工资变量全都有预期的符号吗? 它们是联合地显著的吗? 试加解释。

13.16 本题利用 JTRAIN.RAW 中的数据，以决定在职培训津贴对雇工平均在职培训小时的影响，三年的基本模型是

$$hrsemp_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t + \beta_1 grant_{it} + \beta_2 grant_{i,t-1}$$
$$+ \beta_3 \log(employ_{it}) + a_i + u_{it}$$

439 （i）用一阶差分估计方程，估计中用了多少个厂商？如果每个厂商都有全部变量（特别是 *hrsemp*）的数据，那么三年期间可利用的总观测次数是多少？

（ii）解释 *grant* 的系数并评价它的显著性。

（iii）$grant_{-1}$不显著，你觉得奇怪吗？作出解释。

（iv）平均而言，较大的厂商对它们的雇员培训得多些还是少些？这个培训上的差别有多大？

附录 13A

关于用一阶差分做混合 OLS 的假定

在本附录中，我们仔细表述一阶差分估计量的有关假定。需要做这些假定的理由相当复杂，可参考 Woodridge（1999，第 10 章）。

假定 FD.1

对每个 i，模型都是

$$y_{it} = \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}, \quad t = 1, \cdots, T$$

式中，β_j 为待估计的诸参数；a_i 为非观测效应。

假定 FD.2

我们有一个来自横截面的随机样本。

为了作出下一个假定，我们令 $\mathbf{X}_i$ 表示在第 i 次横截面观测中所有时期的诸解释变量；也就是 $\mathbf{X}_i$ 含有 x_{itj}（$t = 1, \cdots, T$；$j = 1, \cdots, k$）。这样的表示法是有用的。

假定 FD.3

对每个 t，给定所有时期的诸解释变量和非观测效应，特异误差的期望值是零：$E(u_{it} | \mathbf{X}_i, a_i) = 0$。

当假定 FD.3 成立时，我们有时说 x_{itj}是以非观测效应为条件的严格外生变量（strictly exogenous conditional on the unobserved effect）。意思是说，一旦我们控制了 a_i，那么对所有的 s 和 t，x_{isj}和剩余的误差 u_{it}都是不相关的。FD.3 的一个重要含义是 $E(\Delta u_{it} | \mathbf{X}_i) = 0$（$t = 2, \cdots, T$）。

440 **假定 FD.4**

每个解释变量（至少对某个 i）都在时间上有所变化且诸解释变量之间无完全的线性关系。

在这前四个假定之下，一阶差分估计量是无偏的。关键的一个假定是

FD.3，它意味着诸解释变量的严格外生性。在这几个假定下，我们还可以证明，对给定的 T，随着 $N \to \infty$（或在更一般的情况下），FD（一阶差分）估计量具有一致性。

假定 FD.5

以全部解释变量为条件的误差差分的方差是一个常数：

$$\mathrm{Var}(\Delta u_{it} | \mathbf{X}_i) = \sigma^2, \quad t = 2, \cdots, T$$

假定 FD.6

对所有的 $t \neq s$，以全部解释变量为条件的特异误差的（两个）差分 Δu_{it} 和 Δu_{is} 是不相关的：

$$\mathrm{Cov}(\Delta u_{it}, \Delta u_{is} | \mathbf{X}_i) = 0, \quad t \neq s$$

假定 FD.5 保证误差的差分 Δu_{it} 有同方差性。假定 FD.6 则表明差分后的误差没有时序相关性，这就意味着 u_{it} 遵循时间上的一个随机游走模型（见第 11 章）。在假定 FD.1～FD.6 下，β_j 的 FD 估计量是（以诸解释变量为条件的）最优线性无偏估计量。

假定 FD.7

以 $\mathbf{X}_i$ 为条件的 Δu_{it} 是独立且同分布正态随机变量。

当我们增设假定 *FD*.7 时，*FD* 估计量就是正态分布的，而且对差分做混合 *OLS* 所给出的 t 和 F 统计量，有精确的 t 和 F 分布。如果不做假定 *FD*.7，我们就得依赖通常渐近的迫近方法了。

第 14 章　高深的综列数据方法

441 本章中，我们讨论估计非观测效应综列数据模型的两个方法，这些方法的使用至少和一阶差分法一样地普遍。虽然它们颇难以描述和实施，却得到了几个计量经济学软件包的支持。

在14.1节，我们讨论固定效应估计量，它像一阶差分法那样，也在估计之前先做一个变换，把非观测效应 a_i 消除掉，任何不随时间而变的解释变量也将随着 a_i 一道被排除掉。

当我们认为非观测效应与所有的解释变量都不相关时，随机效应估计量更具吸引力。如果在方程中有了良好的控制，那么我们也许会相信，剩下的未被考虑的异因（heterogeneity）只能引起复合误差项中的时序相关，而不会产生复合误差与诸解释变量之间的相关。用广义最小二乘法估计随机效应模型是相当容易的，许多计量经济学软件包都能例行处理。

在14.3节，我们将阐明综列数据方法，怎样能用于其他数据结构，包括配对（matched pairs）和聚类样本（cluster samples）的情形。

14.1　固定效应估计法

取一阶差分仅是消除固定效应 a_i 的许多方法之一。在某些假定下，起到

更好作用的另一种方法，是所谓**固定效应变换**（fixed effects transformation）。为说明这是怎样的一种方法，且考虑仅有一个解释变量的模型：对每个 i，有

$$y_{it}=\beta_1x_{it}+a_i+u_{it},\quad t=1,2,\cdots,T \tag{14.1}$$

现在对每个 i 求方程在时间上的平均，便得到

$$\bar{y}_i=\beta_1\bar{x}_i+a_i+\bar{u}_i \tag{14.2}$$

式中，$\bar{y}_i=T^{-1}\sum_{t=1}^{T}y_{it}$，如此等等。因 a_i 在时间上固定不变，故它兼出现在
442 式（14.1）和式（14.2）中。如果对每个 i 将式（14.2）从式（14.1）减去，我们便得到

$$y_{it}-\bar{y}_i=\beta_1(x_{it}-\bar{x}_i)+u_{it}-\bar{u}_i,\quad t=1,2,\cdots,T$$

或

$$\ddot{y}_{it}=\beta_1\ddot{x}_{it}+\ddot{u}_{it},\quad t=1,2,\cdots,T \tag{14.3}$$

其中，$\ddot{y}_{it}=y_{it}-\bar{y}_i$ 是 y 的**除去时间均值后的数据**（time-demeaned data）；对 $\ddot{x}_{it}$ 和 $\ddot{u}_{it}$ 的解释是类似的。固定效应变换又称**组内变换**（within transformation）。方程（14.3）的要点在于非观测效应 a_i 已随之消失，从而示意我们用混合 OLS 去估计式（14.3）。基于除时间均值变量的混合 OLS 估计量被称为**固定效应估计量**（fixed effects estimator）或**组内估计量**（within estimator）。后一种称谓是因为用于式（14.3）的 OLS 使用了 y 和 x 在每一横截面观测之内的时间变异（time variation）。

当我们对横截面方程（14.2）使用 OLS 估计量时，就给出了**组间估计量**（between estimator）：同时用 y 和 x 的时间平均代替 y 和 x，然后做一个横截面回归。我们不打算详细论述组间估计量，因为如果 a_i 与 $\bar{x}_i$ 相关，它将是有偏误的（见习题 14.2）。而如果我们认为 a_i 与 x_{it} 不相关，则使用**随机效应估计量**（random effects estimator）要更好些，这是我们在 14.2 节将要讨论的。组间估计量忽视了变量怎样在时间上变化所提供的重要信息。

把更多的解释变量加进到方程中来，不会引起什么变化。原始模型将是

$$y_{it}=\beta_1x_{it1}+\beta_2x_{it2}+\cdots+\beta_kx_{itk}+a_i+u_{it},\quad t=1,2,\cdots,T \tag{14.4}$$

我们只须对每一个解释变量——包括诸如时期虚拟变量——除去其时间均值，然后利用全部除时间均值的变量做混合 OLS 回归即可。对每个 i 的一般除去时间平均后的方程是

$$\ddot{y}_{it}=\beta_1\ddot{x}_{it1}+\beta_2\ddot{x}_{it2}+\cdots+\beta_k\ddot{x}_{itk}+\ddot{u}_{it},\quad t=1,2,\cdots,T \tag{14.5}$$

我们用混合 OLS 来估计它。

443 在对诸解释变量做严格外生性假定下，固定效用估计量是无偏的：粗略地说，特异误差 u_{it} 应与所有时期的每一个解释变量都不相关。（关于假定的准确表达，见于章末附录。）固定效应估计量正好比用于估计一阶差分方程的情形那样，容许 a_i 与任何时期的解释变量有任意的相关关系。正因为如

此，凡是在时间上恒定的解释变量都必定随固定效应变换而消失：如果 x_{it} 在时间 t 上是恒定的，则对所有的 i 和 t，$x_{it}=0$。因此，我们不能把诸如性别或某城市是否靠近河流等变量引进来。

问题 14.1 假如在一个家庭储蓄方程中，我们令 $kids_{it}$ 表示第 i 个家庭在第 t 年拥有的儿童人数，$t=1990$ 年、1991 年和 1992 年。如果儿童人数在这 3 年期间对样本中的大多数家庭来说都没有变化，那么，为了估计儿童人数对储蓄的影响，会出现些什么问题呢？

为使简单的 OLS 分析有效，还需要做的假定是误差 u_{it} 的同方差性和无时序相关性：参看本章附录。

还有一个不明显的问题是，怎样决定固定效应估计量的自由度。当我们用混合 OLS 估计除时间均值的方程（14.5）时，我们总共有 NT 个观测值和 k 个自变量。[注意式（14.5）中没有截距，它被固定效应变换消去了。] 因此表面看来，我们有（$NT-k$）个自由度，但这样的计算是不对的。对每个横截面观测 i，由于取时间平均我们失去一个自由度。换言之，对每个 i，除均值后的误差 u_{it} 将对时间 t 加总到零，以致损失一个自由度。因此，适当的自由度是 $df=NT-N-k=N(T-1)-k$。幸而，凡是有固定效应估计功能的现代回归软件包都能适当地算出 df。但是，如果我们需要自己来做除时间均值的运算并且用混合 OLS 进行估计，我们就有必要对标准误和 t 统计量加以校正。

例 14.1 在职培训对厂商报废率的影响

我们利用 54 个厂商在 1987 年、1988 年和 1989 年所报道的每年报废率数据。在 1988 年以前没有任何厂商享受到培训津贴；到 1988 年有 19 个厂商获得津贴；而在 1989 年又有 10 个不同的厂商获得津贴。因此，我们还须考虑 1988 年受到额外在职培训的工人在 1989 年具有更高生产力的可能性。这可通过引进一个津贴标示变量的滞后值而轻易地达到目的。还可引进 1988 年和 1989 年的年虚拟变量。结果见于表 14.1。

表 14.1 **报废率方程的固定效应估计**

	因变量：log(*scrap*)
自变量	
*d*88	−0.080 (0.109)
*d*89	−0.247 (0.133)
grant	−0.252 (0.151)

$grant_{-1}$	-0.422 (0.210)
观测次数	162
自由度	104
R-平方	0.201

444 我们报道结果所采取的方式是要强调必须根据非观测效应模型（14.4）来解释那些估计值。我们明显地控制着 a_i 中的不随时间而变的效应。除掉时间均值虽然能让我们较好估计诸 β_j，但式（14.5）并不是解释估计值的最好方程。

有趣的是，培训津贴滞后效应的估计值远远大于当年的效应：在职培训至少有一年以后的效应。由于因变量取了对数形式，所以预料 1988 年获得津贴的厂商将于 1989 年降低其报废率的 34.4%［$\exp(-0.422)-1\approx-0.344$］；$grant_{-1}$ 的系数相对于双侧假设是在 5% 水平上显著的。$grant$ 的系数则在 10% 水平上显著且系数的大小并非无足轻重。注意，df 算出为 $N(T-1)-k=54(3-1)-4=104$。

$d89$ 的系数表明，即令在职培训津贴没有出现，1989 年的报废率远低于基年 1987 年。因此，考虑好这些总效应是重要的。如果我们忽略了年虚拟变量，就会把劳动生产力的长期增长量归于在职培训津贴。表 14.1 表明，即使控制了生产力的总趋势，在职培训津贴仍估计有一个大的效应。

最后，考虑模型中的滞后效应是很关键的。如果我们忽视了 $grant_{-1}$，那就等于假定在职培训的效应不会延续到下一年。当我们去掉 $grant_{-1}$ 时，对 $grant$ 的估计将是 -0.082（$t=-0.65$）；这比原来的小得多且统计上不显著。

问题 14.2

根据密歇根计划，如果某厂商在某年领取了津贴，它就没有资格在下一年再领取津贴。这对于 $grant$ 与 $grant_{-1}$ 之间的相关意味着什么？

当我们用固定效应方法来估计非观测效应模型时，究竟应该怎样衡量拟合优度，我们还不清楚。表 14.1 所给的 R-平方是根据组内变换计算的：它是从式（14.5）的估计中得到的 R-平方。因此，应把它解释为 y_{it} 的时间变异被诸解释变量的时间变异解释了的部分。关于 R-平方的计算，还可以有其他方法，以下我们讨论其中的一个方法。

虽然不能把时间上恒常的那些变量本身包括到固定效应模型中来，却能把它们同随时间而变的变量，特别是年虚拟变量交互起来分析。例如，在一个工资方程中，对我们的样本中的每一个人来说，教育在时间上是不变的，我们就可以把教育同每一个年虚拟变量交互起来，看教育的回报是怎样在时

间上发生了变化的。但我们无法利用固定效应方法去估计在基期的教育回报——这意味着，我们无法估计任一个时期的教育回报——只能看到每年的教育回报是怎样不同于基期的。

当我们把全部年虚拟变量都包括进来时——即除第一年外每年都安排有一个年虚拟变量——那么，对于任何在时间上的变化为一常数的变量，我们都无法估计其效应。一个例子是，假如在一个样本中，每人每年都在工作，因此每过一年每人的工作经验就增长一年。现在，考虑这样的以年计的工作
445 经验综列数据。a_i 的出现说明了在初始时期不同的人有不同的工作经验（以年计），但从此以后，每增加一年工作经验的效应就无法同总的时间效应区分开来了（因为每人都增长了同样多的工作经验）。如果不分别安排这些年虚拟变量，而是考虑一个线性时间趋势，这一结论也是正确的：因为对每个人来说，经验都无法区别于线性趋势。

例 14.2　在一段时间里的教育回报发生了变化吗？

所用的 WAGPAN.RAW 数据取自维拉和弗比克（Vella and Verbeek, 1998）。样本中的 545 个男子在 1980—1987 年期间，每人每年都在工作。在这段时间里，数据集里的一些变量随时间而改变的，主要有工作经验、婚姻情况及工会会员身份这三个重要变量。另一些变量，诸如种族和教育等重要变量，则不因时间而变。如果我们使用固定效应（或一阶差分）方法，就不能把种族、教育或经验放到方程中来。然而我们能把教育与所有从 1981—1987 年的年虚拟变量的交互作用放进来，以便检验在这段时间里教育回报是否不变。我们用 $\log(wage)$ 作为因变量，还用了经验的一个二次式，婚否、是否工会会员等虚拟变量，全套年虚拟变量，以及交互作用项 $d81\cdot educ$，$d82\cdot educ$，…，$d87\cdot educ$（作为自变量）。

对所有这些交互作用项的估计值都是正的，并且一般地看，越是近年，估计值越大。$d87\cdot educ$ 有最大的系数 0.030，其 $t=2.48$。就是说，估计 1987 年的教育回报比 1980 年基年的要多 3 个百分点。（由于上述原因，我们估计不到基年的教育回报。）其他显著的交互作用项是 $d86\cdot educ$（系数 $=0.027, t=2.23$）。对更早年的估计值更小，且在 5% 水平上相对于双侧对立假设来说不显著。如果我们对全部七个交互项做一个联合 F 显著性检验，将得到 p 值 $=0.28$。这里给出一个例子：一组变量中虽有某些变量个别而论是显著的，但全组联合起来看，却是不显著的。[F 检验的 df 是 7 和 3 799；后一数值来自 $N(T-1)-k=545(8-1)-16=3\ 799$。] 一般地说，这些结果符合教育回报在这一时期的增长。

虚拟变量回归

关于固定效应模型，传统的观点认为，非观测效应 a_i 对每个 i 来说都

是有待估计的一个参数。于是，方程（14.4）中的 a_i 就是第 i 个人（或第 i 个厂商，第 i 个城市，等等）的截距，它和诸 β_j 一起有待于我们去估计。（显然，如果只有一个横截面，这是做不到的：这里只有 N 个观测值。）对每个 i 估计一个截矩的方法，是连同诸解释变量在一起，给每一个横截面观测（单元）安排一个虚拟变量（也许还给每个时期安排有虚拟变量）。这一方法常被称做**虚拟变量回归**（dummy variable regression）。即使 N 还不是很大（比方在例 14.1 中 $N=54$）时，使用此法的结果都会造成许多的解释变量，以致在大多数情况下解释变量多到无法表明如何完成对回归的估计。因此，虚拟变量法对含有许多横截面观测（单元）的综列数据集来说是不很现实的。 446

然而，虚拟变量回归有一些令人感兴趣的特点。最主要的是，它所给出的 β_j 的估计值，和我们对除均值后的数据所做回归时将会得到的恰好一样，而且标准误和其他主要统计量也是雷同的。因此，固定效应估计量可以从虚拟变量回归得到。做虚拟变量回归的一个好处是，可以直接地算出恰当的自由度。但由于现在许多计量经济学软件包都备有编好程序的固定效应法供人选用，这一好处已退居次要了。

从虚拟变量回归算出的 R-平方通常都比较高。这是因为我们对每一横截面单元都安排了一个虚拟变量进来，以致能解释数据中的变异的大部分。例如，我们利用虚拟变量回归（由于 $N=22$，这是可能的），通过固定效应估计例 13.8 中的非观测效应模型，便得到 $R^2=0.933$。我们不应对这个大的 R-平方值过于兴奋：兼用了年和城市虚拟变量所以能解释失业贴补中的大部分变异，就没有什么可惊奇的了。恰如例 13.8 中的情形，对 EZ 虚拟变量的估计值比 R^2 更重要。

从虚拟变量回归得到的 R-平方，可按通常的方法，用来计算 F 检验。当然，这里假定了经典线性模型是成立的（见章末附录）。特别是，我们能够检验所有的横截面虚拟变量（共 $N-1$ 个，因为要选择 1 个单元作为基组）的联合显著性。无约束 R-平方得自于含有全部虚拟变量的一个回归；受约束的 R-平方则来自不含有这些虚拟变量的回归。在绝大多数的应用研究中，这些虚拟变量都会是联合地显著的。

有时，估计的截距 $\hat{a}_i$ 是人们所关注的。这种情形出现在人们想研究 $\hat{a}_i$ 是怎样在 i 中分布的时候，或人们要检查某特定厂商或城市的 $\hat{a}_i$ 是否高于或低于样本平均值。这些估计值虽然可直接从虚拟变量回归得到，但备有固定效应程序的软件包却很少把它们报告出来（因为这些估计值 $\hat{a}_i$ 为数太多了）。且不管 N 多大，在做了固定效应估计之后，要计算 $\hat{a}_i$ 是相当容易的：

$$\hat{a}_i=\bar{y}_i-\hat{\beta}_1\bar{x}_{i1}-\cdots-\hat{\beta}_k\bar{x}_{ik},\quad i=1,2,\cdots,N \tag{14.6}$$

式中，字母上方横线指对时间的平均；而 $\hat{\beta}_j$ 是固定效应估计量。例如，在控制各种因时而变的因素下，估计了一个犯罪模型，我们就能求出某一城市的 $\hat{a}_i$，借以判断其导致犯罪的非观测效应是高于还是低于平均水平。

在大多数研究中，$\hat{\beta}_j$ 才是人们的兴趣所在，因而除去时间平均后的方

程被用来求 β_j 的估计值。另外，一般地说，最好把 a_i 看做在组内变换过程中，我们所要控制的遗漏变量。a_i 之所以被估计的理由通常是脆弱的：事实上，即使 $\hat{a}_i$ 是无偏的（在章末附录中的假定 FE.1～FE.4 下），它对给定的 T 随着 $N \rightarrow \infty$ 却是不一致的。理由是因为我们每增加一个横截面观测（单元），也增加了一个新的 a_i。当 T 被固定了，我们就没有对每个 a_i 积累任何信息。随着 T 变得更大，我们可估计 a_i 的更好的估计值，但大多数综列数据集都属于 N 大而 T 小的类型。

是固定效应(FE) 还是一阶差分(FD) ？

447 至今，我们看到了估计非观测效应模型的两种方法：一种是取数据的差分；另一种是除去对时间的平均。我们怎样知道用哪一种好呢？

我们可以立即消除其中一种情况：当 $T=2$ 时，FE 和 FD 两估计值及其全部检验统计量完全一样，故可随便选用一种。取一阶差分有一个好处：几乎不管用的是什么计量经济学软件包，一阶差分法都是直截了当、一目了然的，而且易于计算对异方差性稳健的统计量。

当 $N \geqslant 3$ 时，FE 和 FD 两估计量便不相同了，因为在假定 FE.1 到 FE.4 下，两者又都是一致性的（固定 T，而 $N \rightarrow \infty$）。对于大的 N 和小的 T，FE 和 FD 之间的选择关键在其估计量的相对效率，而这将由特异误差 u_{it} 中的时序相关性来决定。（因效率比较要建立在同方差的误差基础上，故我们做了 u_{it} 的同方差性假定。）

当 u_{it} 无时序相关时，固定效应法比一阶差分更有效（并且得自固定效应的标准误是真实的）。因为固定效应模型的表述几乎总是伴随有时序不相关的特异误差，所以 FE 估计量的使用更为常见。但我们应记住这一假定可能不真实。在许多应用中可以指望随时间而变的非观测因素是时序相关的。如果 u_{it} 遵循一个随机游走——就是说有一个很强的正的时序相关——那么差分 Δu_{it} 是时序无关的，这时一阶差分法是较好的。在许多情况中，u_{it} 表现有某种正的时序相关，却未必达到一个随机游走的程度，这时要比较 FE 和 FD 两估计量的效率就不那么容易。

在做了 FE 估计之后再检验 u_{it} 是否时序无关是困难的：我们能够估计的是除时间平均的误差 $\ddot{u}_{it}$，而不是 u_{it}。但是在 13.3 节中我们讲过怎样去检验误差差分 Δu_{it} 是否时序无关。所以对于这种情形，可用 FD。如果发现在 Δu_{it} 中有较大的负的时序相关，FE 也许是较好的方法。兼使用两种方法常常是一种好主意：如果结果都差不多，那也就无所谓了。

当 T 很大时，尤其是当 N 还不是很大时（比如，$N=20$ 而 $T=30$），使用固定效用估计量必须保持警惕。虽然在经典固定效应假定下对任何 N 和 T 精确分布结果都是适用的，但当 N 小而 T 大时，这些结果对假定情况的违背是极其敏感的。特别是，如果我们遇到单位根过程——参看第 11 章——就可能出现谬误回归的问题。如我们在第 11 章中所看到的，求一个自

积过程的差分将导致一个弱相依过程，从而我们必须引用中心极限的近似方法。在这种情形中，用差分法较好。另一方面，固定效应对违背严格外生性假定的情况却不那么敏感，尤其当 T 较大时。一些作者甚至推荐在出现滞后因变量时（这时显然违背章末附录中的假定 FE.3）估计固定效应模型为好。当过程在时间上是弱相依而 T 较大时，固定效应估计量的偏误可能较小［例如，参阅 Wooldridge（1999，第 11 章）］。

448 当 FE 和 FD 给出大不相同的结果时，如何在两者之间作出取舍是困难的。同时报告两组结果并试图确定差异的原因所在，并非没有意义。

非平衡综列数据的固定效应法

一些综列数据集，特别是关于个人或厂商的，样本中缺少了某些横截面单元的某年数据，这至少是常见的现象。这时，我们称数据集为**非平衡综列数据**（unbalanced panel）。如何操作非平衡综列数据的固定效应法，并不比操作平衡综列数据时的情况困难多少。设 T_i 为横截面单元 i 的时期个数，我们只须用 T_i 个观测值去做除时间平均的运算。观测值的总数将是 $T_1+T_2+\cdots+T_N$。和平衡综列数据的情形一样，对每一个横截面观测（单元）都因除时间平均运算而失去一个自由度。任何一个做固定效应的回归软件包都会作出适当的自由度损失调整。虚拟变量回归所经历的程序也和平衡综列数据情形完全一样，但要适当地调整自由度。

容易看出，在固定效应分析中，只有一个时期的那些单元将不起任何作用。对于这样的观测单元，除时间平均的计算结果都是零，自然无补于估计。（如果对所有的 i，T_i 最多是 2，我们就可以用一阶差方法：如果某个 i 的 $T_i=1$，我们就无从取两时期的差分。）

要明确综列数据为什么会变成非平衡的，则属于（非平衡综列数据）比较困难的问题。例如，对城市或州来说，有时一些重要变量的某某年数据缺失了。如果是一些 i 之所以缺失数据的理由与特异误差 u_{it} 不相关，非平衡综列数据就不会引发什么问题。当我们的数据是关于个人、家庭或厂商的，事情往往比较复杂。试想，比方说，我们在 1990 年抽取了制造业厂商的一个随机样本，并且我们的兴趣在于检验工会化是怎样影响厂商的获利性的。我们希望能利用综列数据分析去控制一些非观测的劳工与经理特性；这些特性不但影响着获利性，还可能与参加工会的人占全厂劳工的比例有相关关系。如果我们在随后的一些年里再次收集数据，有些厂商会因企业倒闭或被其他公司兼并而消失。如果是这样，我们在随后的时期所获得的也许是一个非随机样本了。问题是，如果我们把固定效应法应用到这个非平衡综列数据上，怎样才会得到无偏（或至少是一致）估计量呢？

如果一个厂商离开样本［称耗损（attrition）］的理由与特异误差——指那些随时间而变且影响着利润的误差——相关，那么，由此造成的样本截面问题（见第 9 章）就会导致偏误的估计量。这是本例中的一个严肃的考虑。

然而，固定效应分析有一个可利用的地方，就是它容许耗损与非观测效应 a_i 相关。意思是说，初始抽样中的一些单元比另一些更有可能从调查中消失，而这一点已被 a_i 描述了。

例 14.3　在职培训对厂商报废率的影响

我们把两个变量加进到表 14.1 的分析中：$\log(sales_{it})$和 $\log(employ_{it})$，其中 *sales* 为每年厂商销售量；而 *employ* 为雇工人数。在 54 个厂商中有 3 个因为没有销售和就业数据，所以完全从分析中消失。另外有 5 个观测值由于在某些年里缺失了一个或两个变量的数据而消失，这样就剩下 $n = 148$。把固定效应法用到这个非平衡综合数据上并不改变基本的故事情节，只是津贴效应变大了一些： 449

$$\hat{\beta}_{grant} = -0.297, t_{grant} = -1.89; \hat{\beta}_{grant_{-1}} = -0.536, t_{grant_{-1}} = -2.389$$

解决综合数据中的耗损问题是复杂的，而且超出了本书的范围。[可参阅 Wooldridge（1999，第 17 章）]

14.2　随机效应模型

如前，我们从同一个非观测效应模型开始：

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it} \tag{14.7}$$

式中，显式地引进一个截距，使得我们能够假定非观测效应 a_i 有零的值（而无碍于一般性）。通常我们还会考虑和诸解释变量一道出现的时间虚拟变量。不管使用固定效应法还是一阶差分法，目的都是要把 a_i 消去。这是因为 a_i 被认为是与 x_{itj}中之一或多个相关。但设想我们认为 a_i 与任何一个解释变量在任何时期都不相关？那么，通过变换把 a_i 消去就会导致非有效（或效率低）的估计量。

如果假定非观测效应 a_i 与每一个解释变量都不相关：

$$\mathrm{Cov}(x_{itj}, a_i) = 0, \quad t = 1,2,\cdots,T; j = 1,2,\cdots,k \tag{14.8}$$

则方程(14.7)变为一个**随机效应模型**(random effects model)。事实上，理想的随机效应假定包括全部固定效应假定，再加上 a_i 独立于所有时期的每一个解释变量的假定。(关于实际用到的假定，参阅章末附录。)如果我们认为非观测效应 a_i 会与任一个解释变量相关，那就应该对固定效应做一阶差分。

在连同式（14.8）的随机效应假定下，我们应该怎样去估计诸 β_j 呢？如果我们相信 a_i 与解释变量不相关，则可用单一个横截面对 β_j 做一致性估计：根本不需要什么综列数据。认识清楚这一点是重要的。但是使用单一个

横截面显然忽视了其他时期许多有用的信息。我们可以按照混合 OLS 程序利用这些信息，也就是将 y_{it} 对诸解释变量也许还加上时间虚拟变量做 OLS 回归，在随机效应假定下，这样做也能产生 β_j 的一致估计量。但它忽略了模型的基本要点。如果我们定义**复合误差项**（composite error term）为 $v_{it} = a_i + u_{it}$，则式（14.7）可写为

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + v_{it} \tag{14.9}$$

450 由于 a_i 在每个时期都是复合误差的一部分，v_{it} 在不同时间上就应是时序相关的。事实上，在随机效应假定下，有

$$\mathrm{Cov}(v_{it}, v_{is}) = \sigma_a^2/(\sigma_a^2 + \sigma_u^2), \quad t \neq s$$

式中，$\sigma_a^2 = \mathrm{Var}(a_i)$，$\sigma_u^2 = \mathrm{Var}(u_{it})$。误差项中的这一（必然是）正的时序相关可能很大；由于通常的混合 OLS 标准误忽视了这种相关，以致标准误是不正确的，从而通常用的检验统计量也不正确。在第 12 章中我们曾说明怎样可用广义最小二乘法来估计带有（自回归）时序相关的模型，我们也可以用 GLS 解决这里的时序相关性问题。为了使该程序具有良好的性质，我们必须有大的 N 和相对小的 T。我们还假定有一个平衡的综列数据集，尽管我们的方法可推广到非平衡的情形。

为了推导 GLS 变换以消去误差中的时序相关，要用到复杂的矩阵代数［例如参看 Wooldridge（1999，第 10 章）］。但变换本身却是简单的，定义

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{\frac{1}{2}} \tag{14.10}$$

这是一个 0～1 之间的分数。于是，变换方程变为

$$\begin{aligned} y_{it} - \lambda \bar{y}_i = {} & \beta_0(1 - \lambda) + \beta_1(x_{it1} - \lambda \bar{x}_{i1}) + \cdots + \beta_k(x_{itk} - \lambda \bar{x}_{ik}) \\ & + (v_{it} - \lambda \bar{v}_i) \end{aligned} \tag{14.11}$$

式中，字母上方横线表示时间平均。这个方程颇为有趣，它涉及每个变量的一组**准除平均后的数据**（quasi-demeaned data）。固定效应估计量从相应的变量减去其时间平均，而随机效应变换只减去其时间平均的一个分数，这个分数依赖于 σ_u^2，σ_a^2 和时期的个数 T。GLS 估计量不外是方程（14.11）的混合 OLS 估计量，很难看出式（14.11）中的误差是时序不相关的，然而事实却如此。

式（14.11）中的变换容许我们考虑不随时间而变的解释变量，这是随机效应（RE）和固定效应或一阶差分相比的一个优点。［可比较式（14.11）和式（14.3）或式（14.5）——译者注］之所以能够有这一优点，是因为 RE 假定了非观测效应与所有的解释变量都不相关，不管这些解释变量是否随时间而变化。例如，在一个工资方程中，我们可以引进一个即令是不因时间而变的教育变量。但是我们是在假定教育与包含着能力和家庭背景的 a_i 不相关。在许多应用中，要使用综列数据的全部原因，却是为了容许非观测效应与解释变量之间有相关关系。

实际上，参数 λ 是永远不知道的，但又总是可以估计的。有不同的估计方法，例如，可根据混合 OLS 或固定效应做出估计。通常，$\hat{\lambda}$ 采取 $\hat{\lambda} = 1 - \{1/$

$[1+T(\hat{\sigma}_a^2/\hat{\sigma}_u^2)]\}^{\frac{1}{2}}$的形式，其中 $\hat{\sigma}_a^2$ 是 σ_a^2 的一个一致估计量，而 $\hat{\sigma}_u^2$ 是 σ_u^2 的一个一致估计量。这些估计量是根据混合 OLS 残差或固定效应残差计算的。一种可能性是 $\hat{\sigma}_a^2=[NT(T-1)/2-k]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}$，其中 $\hat{v}_{it}$是用混合
451 OLS 估计式(14.9)的残差。给定 $\hat{\sigma}_a^2$，就可通过 $\hat{\sigma}_u^2=\hat{\sigma}_v^2-\hat{\sigma}_a^2$ 来估计 σ_u^2，其中 $\hat{\sigma}_v^2$ 是从混合 OLS 得到的通常的回归标准误的平方。[参看 Wooldridge(1991，第 10 章)以获得更多的关于这些估计量的讨论。]

有许多计量经济学软件包支持随机效应模型的估计，并自动算出 $\hat{\lambda}$ 的一些形式。用 $\hat{\lambda}$ 代替 λ 的可行 GLS 估计量被称做随机效应估计量。在章末附录的随机效应假定下，该估计量是一致（而不是无偏）的，并且对固定的 T 随 N 增大而渐近于正态分布。当 N 小而 T 大时，RE 估计量的性质大多是未知的，尽管在这种情况下肯定还要用到这种估计量。

方程（14.11）使我们能够把 RE 估计量同混合 OLS 以及固定效应两者联系起来。当 $\lambda=0$ 时，得到混合 OLS；当 $\lambda=1$ 时，得到 FE。实际上，估计值 $\hat{\lambda}$ 永远不会是 0 或 1。但如果 $\hat{\lambda}$ 接近零，RE 估计量就会接近混合 OLS 估计值。当非观测效应 a_i 相对地不重要时（因为相对于 σ_u^2 它有小的方差），就会出现这种情形。但更常见的是 σ_a^2 相对地比 σ_u^2 大，这时 λ 将接近 1。随着 T 增大，$\hat{\lambda}$ 趋于 1，从而使得 RE 和 FE 两种估计非常相似。

例 14.4　使用综列数据的一个工资方程

我们再次使用 WAGEPAN.RAW 中的数据估计一个男性的工资方程。我们用三种方法：混合 OLS、随机效应和固定效应。在头两个方法中我们可以把 *educ* 和种族虚拟变量［黑人（*black*）和西班牙裔（*hispan*）］包括进来，而在固定效应分析中这些变量将消失。随时间而变的变量是 *exper*，*exper*^{2}，*union* 和 *married*。如 14.1 节所讨论的，*exper* 将在 FE 分析中消失（但 *exper*^{2} 会保留下来）。每个回归都包含有全部年虚拟变量。估计结果见表 14.2。

混合 OLS 和随机效应两个估计量给出的 *educ*，*black* 和 *hispan* 的系数是相似的。混合 OLS 标准误就是通常的 OLS 标准误。这些标准误由于忽略了正的时序相关而低估了真正的标准误；我们在这里提及它们，只是为了便于比较。关于经验的变化则有些不同，并且在随机效应估计中，已婚和入会的增益明显地下降了。当我们用固定效应法把非观测效应全部消去后，已婚的增益减到大约 4.7%，尽管它仍是统计上显著的。结婚增益的下降符合越是能干的男子——为一个更高的非观测效应 a_i 所刻画——就越可能是结了婚的那种想法。因此，在混合 OLS 估计中，结婚增益的一大部分反映这样一个事实：结了婚的男子即使不结婚也挣更多的钱。余下的 4.7% 至少有两种可能的解释：（1）结了婚的男子确实有更高的生产力，或者（2）因为成婚是一个稳定性信号，所以雇主给已婚男子某种津贴。我们无法区分这两种假设。

问题 14.3 ☞ 由固定效应法估计的会员增益大约比用 OLS 所估计的低 10 个百分点。这对于 *union* 和非观测效应之间的相关给了我们什么强有力的提示？

表 14.2 **工资方程的三种不同的估计量**

452

因变量：log(*wage*)			
自变量	混合 OLS	随机效应	固定效应
educ	0.091 (0.005)	0.092 (0.011)	—
black	−0.139 (0.024)	−0.139 (0.048)	—
hispan	0.016 (0.021)	0.022 (0.043)	—
exper	0.067 (0.014)	0.106 (0.015)	—
$exper^2$	−0.002 4 (0.000 8)	−0.004 7 (0.000 7)	−0.005 2 (0.000 7)
married	0.108 (0.016)	0.064 (0.017)	0.047 (0.018)
union	0.182 (0.017)	0.106 (0.018)	0.080 (0.019)

在随机效应估计中，λ 的估计值是 $\hat{\lambda}=0.643$，这说明为什么就随时间而变的变量而言，RE 估计值更靠近 FE 而不是混合 OLS 估计值。

随机效应还是固定效应？

在阅读经验性作品时，你会发现作者选择固定抑或随机效应的依据是，最好把 a_i（或作者用的什么别的符号）看做待估的参数呢，抑或看成一个随机变量所表现的结果。当我们不能把观测值当做从一个大总体中随机抽样的结果——例如我们拥有的关于州或郡（县）的数据——时，通常都把 a_i 看做待估的参数，这样我们就得使用固定效应法。请记住，使用固定效应法等同于考虑每个观测（单元）有一个不同的截距，并且可通过引进虚拟变量

或利用式（14.6）去估计它们。

即使我们决定把 a_i 看成随机变量，我们还必须决定 a_i 是否与诸解释变

453 量不相关。人们有时错误地认为，假定了 a_i 是随机的就自动地意味着随机效应法是适当的估计策略。如果我们（确实）能够假定（肯定）a_i 与全部 x_{it} 都不相关，那么随机效应法是适宜的。但若 a_i 与某些解释变量相关，则仍然需要固定效应法（或一次差分法）；如果用了 RE，一般地说估计量将是非一致的。

假定特异误差与解释变量在所有时期都不相关，那么 FE 和 RE 两个估计量的比较可以成为 a_i 与 x_{itj} 是否相关的一种检验。豪斯曼（Hausman，1978）首先提出这种检验。一些计量经济学软件包在章末附录所列举的理想随机效应假定下例行地计算了这个检验统计量。关于该统计量的细节，可参阅 Wooldridge（1999，第 10 章）。

14.3 把综列数据方法用于其他数据结构

差分法、固定效应法和随机效应法可用于不涉及时间的一些数据结构。例如，在人口统计学中常用胞亲（siblings）（有时是孪生配对）去控制观测不到的家庭及其背景特性。对胞亲取差分或者说得更一般，在一个家庭内部作组内变换，以消除可能与解释变量相关的家庭效应。

作为一个例子，杰罗尼姆斯和考伦曼（Geronimus and Korenman，1992）利用多双姐妹数据，研究十几岁的童年生育对未来造成的经济后果。以相对于需要——诸如相对于抚养小孩的个数——的收入作为后果，模型是

$$\log(incneeds_{fs}) = \beta_0 + \delta_0 sister2_s + \beta_1 teenbrth_{fs} + \beta_2 age_{fs} + other\ factors + a_f + u_{fs} \qquad (14.12)$$

式中，f 为家庭；s 为该家庭中的姐妹之一。姐姐的截距是 β_0，妹妹的截距是 $\beta_0+\delta_0$。所关注的变量是 $teenbrth_{fs}$，这是一个二值变量，当家庭 f 的姐妹 s 在十几岁时生育小孩，它就等于 1。变量 age_{fs} 是家庭 f 中姐妹 s 的当时年龄。杰罗尼姆斯和考伦曼还用了一些其他控制变量。仅随家庭而变的非观测效应 a_f 是一个观测不到的家庭效应或家庭固定效应。分析中的主要问题是 $teenbrth$ 与家庭效应是否相关。如果是相关的，那么混合家庭和姐妹的 OLS 分析就会对童年（十几岁）当母亲的经济后果作出一个有偏误的估计。解决这个问题的方法是简单的：在每个家庭之内、在姐妹之间取式（14.12）的差分，于是得

$$\Delta\log(incneeds) = \delta_0 + \beta_1\Delta teenbrth + \beta_2\Delta age + \cdots + \Delta u \qquad (14.13)$$

这样就消除掉了家庭效应 a_f，并可用 OLS 估计这个结果方程。注意这里并没有任何时间因次：差分是在一家之内姐妹之间进行的。

454 杰罗尼姆斯和考伦曼利用1982年在全国青年妇女纵横数据调查中的129对姐妹，先用混合OLS估计β_1为−0.33和−0.26，后一个估计值是对家庭背景变量（诸如父母亲所受教育）作了控制之后得到的；两个估计值都在统计上非常显著［见Geronimus和Korenman（1992）的表3］。因此，童年当母亲对今后家庭收入有相当大的影响。然而，再对差分方程进行估计时，*teenbrth*的系数却变为−0.08，既小而又统计上不显著。这表明大体上说影响未来收入的是一个妇女的家庭背景，而不是童年时代当了妈妈。

问题 14.4 如果用了差分法，那么在式（14.12）中引进父母亲的种族虚拟变量还会有意义吗？

杰罗尼姆斯和考伦曼还调研了一些其他方面的后果和其他的数据组；有些时候，这种家庭之内的估计结果是经济上重要且统计上显著的。他们还指出，当控制了姐妹的教育水平之后，这些效应会怎样全部消失的

阿申费尔特和克鲁格（Ashenfelter and Krueger，1994）利用整套差分方法估计了受教育的回报。他们曾获得一个由149对情况相同的孪生子（identical twins）构成的样本，并收集了关于收入、教育以及其他变量的信息。使用情况相同的孪生子，理由在于他们应有相同的潜在能力。通过孪生子之间的差分而不是对混合数据做OLS就可把这种能力因素差分掉。因为情况相同的孪生子在年龄、性别和种族上都是一样的，这些因素都将从差分方程中消失。因此，阿申费尔特和克鲁格将log(*earnings*)的差分对教育的差分做回归，并估计出教育的回报约为9.2%（$t=3.83$）。有趣的是，这比（控制了性别、年龄和种族的）混合OLS估计值8.4%还要大。阿申费尔特和克鲁格还用随机效应法估计了方程并得到教育回报为8.7%。（参阅他们的论文中的表5。）这个随机效应分析在操作上无异于两时期的综列数据情形。

阿申费尔特和克鲁格（Geronimus and Korenman，1992）以及阿申费尔特和克鲁格（Ashenfelter and Krueger，1994）所用的样本是（所谓）**配对样本**（matched pair samples）的两个例子。一般地说，固定和随机效应法都可用于（所谓）**聚类样本**（cluster sample）。后者指横截面数据集里的每一观测都属于一个明确定义的聚类。在上述例子中，每个家庭是一个聚类。作为另一个例子，假使我们有多个养老金计划的参与数据；每个厂商提供不止一个计划。于是，我们可把每个厂商看做一个聚类，从而相当清楚地看到，观测不到的厂商效应乃是决定厂商内部（组内）养老金计划参与率的一个重要因素。

从多所学校抽样得到的学生教育数据构成一个聚类样本，这里每所学校是一个聚类。因为一个聚类之内的结果很可能是相关的，所以考虑非观测聚类效应就特别重要。当我们认为这种**非观测聚类效应**（unobserved cluster effect）——式（14.12）中的a_i就是一个例子——与解释变量之一或多个相关时，固定效应估计法就是较好的方法。这时，我们可以仅仅引进在聚类内至少是多少有些变化的解释变量。每个聚类的大小很少是一样的，因此常要

用到非平衡综列数据的固定效应法。

455 假如聚类效应与所有解释变量都不相关，那么随机效应法也可用于非平衡聚类。对于这种情形，我们还可以利用混合 OLS，但这时除非聚类之内无相关关系，否则通常的标准误是不正确的。一些回归软件包具有简单的命令去修正标准误和通常的检验统计量去检验一般性的聚类内相关关系（以及异方差性）。这些校正方法和用于综列数据集的、混合 OLS 方法是一样的。见例 13.9 中的报告。例如，帕普克（Papke，1999）估计了一个线性概率模型，这个模型根据厂商是否采取明确制定的捐赠计划，以判断它是否继续执行既定的福利养老金计划。因为这里很可能出现一种厂商效应，致使同一厂商之内的不同计划之间有相关关系，帕普克对线性概率模型中的常用 OLS 标准误作了聚类抽样以及异方差性方面的校正。

▶ 小结

我们学习了估计带有非观测（即观测不到的）效应的综列数据模型的两种常用方法。当特异误差无时序相关（并且具有同方差性）时，固定效应法和一阶差分法相比是更有效的。这里我们并没有假定非观测效应 a_i 与解释变量之间是否相关。和做一阶差分一样，任何不随时间而变的解释变量将从分析中消失。虽然固定效应法直接适用于非平衡综列数据，但我们必须假定某些时期数据缺落的原因与特异误差之间无系统的关系。

当我们认为非观测效应与所有解释变量都不相关时，随机效应估计量是适用的。这时可把 a_i 留在误差项里，并且由此产生的时序相关可通过广义最小二乘估计法加以处理。对准除平均后的数据做混合回归，就能方便地达到做可行广义最小二乘的目的。所估计的变换参数值 $\hat{\lambda}$ 示意我们，究竟估计结果更接近于混合 OLS 估计值呢抑或固定效应估计值？如果全部随机效应假定都成立，随机效应估计量就是渐近地——在固定 T 而随着 N 变大时——比混合 OLS、一阶差分或固定效应估计量（这些全都是无偏的、一致的且渐近于正态的）更有效。

最后，当我们拥有配对或聚类样本时，第 13 章和第 14 两章中所讲的综列数据方法都可利用。通过差分或组内变量消去了聚类效应。如果聚类效应与解释变量无关，就可使用混合 OLS，但这时应对标准误和检验统计量做聚类相关方面的校正。还有可能用到随机效应估计法。

关键术语

聚类效应

准除平均后数据

聚类样本
复合误差项
虚拟变量回归
固定效应估计量
456 固定效应变换
配对样本

随机效应估计量
随机效应模型
除时间平均后数据
非平衡综列数据
组内估计量
组内变换

习　题

14.1　假使式（14.4）中的特异误差 $\{u_{it}: t=1, 2, \cdots, T\}$ 是时序不相关的且有恒定的方差 σ_u^2。证明相邻差分 Δu_{it} 和 $\Delta u_{i,t+1}$ 的相关系数为 -0.5。因此，在理想的 FE 假定下，一阶差分导致一个已知其值的负序列相关。

14.2　当只有一个解释变量时，用来求组间估计量的方程是

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i$$

式中，上方横线表示对时间取平均。因为方程中含有一个截距项，所以不妨假定 $\mathrm{E}(a_i)=0$。假使 $\bar{u}_i$ 与 $\bar{x}_i$ 不相关，但对一切 t（以及 i，这是由于在横截面中进行了随机抽样）有 $\mathrm{Cov}(x_{it}, a_i)=\sigma_{xa}$。

（i）令 $\tilde{\beta}_1$ 为组间估计量，也就是对时间平均做 OLS 的估计量，证明

$$\mathrm{plim}\tilde{\beta}_1 = \beta_1 + \sigma_{xa}/\mathrm{Var}(\bar{x}_i)$$

式中概率极限是对 $N \to \infty$ 而定义的。［提示：参看方程（5.5）和方程（5.6）。］

（ii）再假定对所有 $t=1, 2, \cdots, T$，x_{it} 都是不相关的且有恒定方差 σ_x^2。证明 $\mathrm{plim}\tilde{\beta}_1 = \beta_1 + T(\sigma_{xa}/\sigma_x^2)$。

（iii）如果解释变量在时间上不是非常高度相关的，那么是否时期越多组间估计量的非一致性越小呢？第（ii）部分对这个问题有什么提示？

14.3　在随机效应模型中，定义复合误差为 $v_{it}=a_i+u_{it}$（a_i 与 u_{it} 不相关），而且 u_{it} 有恒定方差 σ_u^2 并且是时序不相关的。定义 $e_{it}=v_{it}-\lambda\bar{v}_i$，其中 λ 由（14.10）给出。证明 e_{it} 有零均值、恒定方差，而且是时序不相关的。

14.4　为确定大学生校际运动成绩对申请人的影响，你对 1 类（Division I）大学的一个样本收集了 1985 年、1990 年和 1995 年的入学申请数据。

（i）你会把怎样度量体育成绩的变量放在方程中吗？你对时期的编排有些什么争议？

（ii）你想在方程中控制些什么别的因素？

（iii）试写出一个方程，用以估计体育成绩对申请人数的百分比变化的影响。你会用什么方法估计方程呢？为什么选用这样的方法？

14.5 设想你能在某一个学期对大学三年级和四年级学生的一个随机样本收集到他们所修的每门功课的如下数据：标准化的期终考试成绩；到课百分率；像征是否学生的主修课的一个虚拟变量；学期开始前的累积平均学习
457 成绩以及 *SAT* 分数。

(i) 你为什么会把这些数据归类为聚类样本？你预期能从一个典型学生得到大概多少个观测值？

(ii) 写出一个类似于方程（14.12）那样的模型，用到课率和其他特征去解释期终考试成绩。以 s 做学生下标和 c 做课程下标，对同一个学生哪个变量是不变的？

(iii) 如果你把所有的数据混合起来并使用 OLS，那么，对影响着成绩和到课率的非观测学生特征，你正在做些什么假定呢？在这方面 *SAT* 和事前 *GPA* 扮演着什么角色呢？

(iv) 如果你认为 *SAT* 和事前 *GPA* 不能适当地表征学生的能力，你会怎样估计到课率对期终考试的影响呢？

计算机习题

14.6 本题使用 RENTAL.RAW 中的数据。数据是关于 1980 年和 1990 年一些大学城的房租和其他变量的，要看看较多的学生到来是否对房租有影响。非观测效应模型是

$$\log(rent_{it}) = \beta_0 + \delta_0 y90_t + \beta_1 \log(pop_{it}) + \beta_2 \log(avginc_{it}) \\ + \beta_3 pctstu_{it} + a_i + u_{it}$$

式中，*pop* 为城市人口；*avginc* 为平均收入；*pctstu* 为学生人口占城市人口的百分比（按学年计）。

(i) 用混合 OLS 估计方程并按标准形式报告结果。你从 1990 年虚拟变量的估计值得到些什么结论？你得到了什么 $\hat{\beta}_{pctstu}$ 值？

(ii) 你在第（i）部分报告的标准误真实吗？作出解释。

(iii) 现在取方程的差分，再用 OLS 去估计。比较 β_{pctstu} 的估计值和第（i）部分所估计的。学生人口的相对大小看来是否对房租有影响？

(iv) 用固定效应估计模型，以证实你得到和第（iii）部分得到的估计值和标准误完全相同。

14.7 本题使用 CRIME4.RAW 中的数据。

(i) 使用固定效用法而不是差分法重新估计例 13.9 中的关于犯罪的非观测效应模型。系数的符号和大小有什么明显的变化？其统计显著性又怎样？

(ii) 在数据集中加进每个工资变量的对数，再用固定效应法估计模型。加进这些变量后，第（i）部分关于判刑公正变量的系数会怎样地受到影响？

(iii) 第（ii）部分的工资变量都带有所预期的符号吗？作出解释。它们
458 是联合地显著的吗？

14.8 本题使用 JTRAIN.RAW 的数据，以决定在职培训津贴对每个雇

工在职培训小时数的影响。三年的基本模型是

$$hrsemp_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t + \beta_1 grant_{it} + \beta_2 grant_{i,t-1} + \beta_3 \log(employ_{it}) + a_i + u_{it}$$

（i）用固定效应法估计方程。在此估计中利用了多少个厂商？如果每个厂商都具有这三年的所有变量的数据（特别是 *hrsemp* 的数据），总观测个数会是多少？

（ii）解释 *grant* 的系数并评论其显著性。

（iii）$grant_{-1}$不显著有什么可奇怪的吗？作出解释。

（iv）平均地说，较大的厂商为它们的职工提供了更多抑或更少的培训？差别有多大？（比方说，职工多 10％的厂商，培训的平均小时数增多或减少了多少？）

14.9　在例 13.8 中，我们用了帕普克（Papke，1994）的失业补贴数据去估计企业特区对失业补贴的影响。帕普克还用过一个能使得每个城市看到自己的时间趋势的模型：

$$\log(uclms_{it}) = a_i + c_i t + \beta_1 ez_{it} + u_{it}$$

式中，a_i 和c_i 都是非观测效应，这样就可考虑城市之间的更多的差异性。

（i）说明如果对上列方程取差分就会得到

$$\Delta\log(uclms_{it}) = c_i + \beta_1 \Delta ez_{it} + \Delta u_{it},\quad t = 2, \cdots, T$$

注意在此差分方程中有一固定效应 c_i。

（ii）用固定效应法估计差分方程。β_1 的估计值是什么？它和例 13.8 中所估计的有很大的差别吗？企业特区的作用仍是统计上显著的吗？

（iii）在第（ii）部分的估计中把全部年虚拟变量加进来，β_1 的估计值出现了什么变化？

14.10　（i）在例 14.4 的工资方程中，为了估计会员工资的增益，职业虚拟变量可能是重要的遗漏变量，解释为什么。

（ii）使用 WAGEPAN.RAW 数据时，把其中的 8 个职业虚拟变量包括进来，再用固定效应法估计方程。*union* 的系数是否改变很大？其统计显著性如何？

14.11　把交互作用项 $union_{it} \cdot t$ 加到表 14.2 所估计的方程中，看看工资的增长是否与会员身份有关。兼用随机和固定效应两种方法估计方程并比较其结果。

附录 14A

关于固定效应和随机效应的假定

459 在本附录中，我们阐述有关固定和随机效应估计法的诸假定。我们还讨论在不同的假定集下，这些估计量的性质。有关性质的证明是相当复杂的，但可参阅 Wooldridge（1999，第 10 章）。

假定 FE.1

对每个 i，模型为

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + a_i + u_{it}, \quad t = 1, 2, \cdots, T$$

式中，β_j 为待估参数。

假定 FE.2

我们有一个横截面（维数）中的随机样本。

假定 FE.3

对每个 t，在给定所有时期的解释变量和非观测效应的条件下，特异误差的期望值是零：

$$E\ (u_{it} \mid \boldsymbol{X}_i,\ a_i)\ = 0$$

假定 FE.4

每个解释变量（至少对某些 i 来说）在时间上有所变化，并且在诸解释变量之间无完全的线性关系。

在这 4 个假定——也就是对一阶差分估计量所作的完全一样的假定——下，固定效用估计量是无偏的。关键的假定仍然是严格外生性假定 FE.3。在同样的这些假定下，对于固定的 T，随着 $N \rightarrow \infty$，FE 估计量是一致的。

假定 FE.5

$$\mathrm{Var}(u_{it} \mid \boldsymbol{X}_i, a_i) = \mathrm{Var}(u_{it}) = \sigma_u^2, \quad t = 1, 2, \cdots, T$$

假定 FE.6

对所有 $t \neq s$，特异误差是不相关的（以所有的解释变量和 a_i 为条件）：

$$\mathrm{Cov}(u_{it}, u_{is} \mid \boldsymbol{X}_i, a_i) = 0$$

在假定 FE.1～FE.6 下，β_j 的固定效应估计量是最优线性无偏估计量。由于 FD 估计量是线性且无偏的，它必然不比 FE 估计量好。使得 FE 比 FD 更好的假定是 FE.6，它意味着特异误差是时序不相关的。

假定 FE.7

460 以 $\boldsymbol{X}_i$ 和 a_i 为条件，u_{it} 是独立的并且是按正态 Normal（0，σ_u^2）同分布的。

假定 FE.7 蕴涵假定 FE.3，FE.5 和 FE.6，但由于假定了特异误差的正态分布而相对更强。如果加上 FE.7，那么 FE 估计量就是正态分布的，从而 t 和 F 统计量都有精确的 t 和 F 分布。如果没有 FE.7，则可借助于渐近迫近法。但是，除非作出特殊的假定，这些迫近方法都要求有大的 N 和小的 T。

理想的随机效应假定包括 FE.1，FE.2，FE.3，FE.5 和 FE.6。这样就能考虑不随时间而变的变量。（虽然可以把 FE.7 加进来，但实际上帮不了我们什么忙。）然而，我们需要增加一些假定，以说明 a_i 是怎样与诸解释变量发生关系的。于是，我们把第 3 个假定加强如下：

假定 RE.3

除 FE.3 外，在给定全部解释变量的条件下，a_i 的期望值是零：

$$\mathrm{E}\ (a_i \mid \boldsymbol{X}_i)\ = 0$$

这个假定是用来排除非观测效应与解释变量之间有任何相关关系的。因为 RE 变换没有完全消除掉时间上的平均，所以我们能够考虑对任何 i 在时间上保持不变的解释变量。

假定 RE.4

在诸解释变量之间不存在完全的线性关系。

我们还需要对 a_i 作如下的同方差性约束：

假定 FE.5

除 FE.5 之外，在给定所有解释变量的条件下，a_i 的方差是个常数：

$$\mathrm{Var}(a_i \mid \boldsymbol{X}_i) = \sigma_a^2$$

在这 6 个随机效应假定（FE.1，FE.2，RE.3，RE.4，RE.5 和 FE.6）之下，随机效应估计量对固定的 T 和不断增大的 N 来说是一致的。实际上，一致性仅需要前 4 个假定。除非我们知道 λ（而不是靠估计而得知的），RE 估计量并非无偏。当 N 很大时，RE 估计量还是渐近正态分布的，并且从准除均值回归（quasi-demeaned regression）得来的、通常的标准误以及 t 和 F 统计量是适用的。[更详细的报导，见于 Wooldridge（1999，第 10 章）。]

第 15 章　工具变量估计与两阶段最小二乘法

461　在本章中，我们进一步研究多元回归模型中的**内生解释变量**（endogenous explanatory variable）问题。在第 3 章中，我们推导出遗漏一个重要变量时 OLS 估计量的偏误；在第 5 章中，我们说明了在**遗漏变量**（omitted variable）的情况下，OLS 通常是非一致性的。第 9 章则证明了，对未观测到的解释变量给出适宜的代理变量，能消除（或至少减轻）遗漏变量偏误。不幸的是，我们不是总能得到适宜的代理变量。

在前两章中，我们解释了存在不随时间变化的遗漏变量的情况下，对综列数据如何用固定效应估计或一阶差分来估计随时间变化的自变量的影响。尽管这些方法非常有用，可我们不是总能获得综列数据的。即使能获得，如果我们的兴趣在于变量的影响，而该变量不随时间变化，它对于我们也几无用处：一阶差分或固定效应估计排除了不随时间变化的变量。此外，迄今为止我们已研究出的综列数据法还不能解决与解释变量相关的随时间而变化的遗漏变量的问题。

在本章中，我们对内生性问题采用了一个不同的方法。你将看到如何用工具变量法（IV）来解决一个或多个解释变量的内生性问题。就应用计量经济学中线性方程的估计而言，两阶段最小二乘法（2SLS 或 TSLS）是第二受人欢迎的，仅次于普通最小二乘。

我们一开始先说明，在存在遗漏变量的情况下，如何用 IV 法来获得一致

性估计量。此外，IV 能用于解决含误差变量（errors-in-variable）的问题，至少是在某些假定下。下一章将证明运用 IV 法如何估计联立方程模型。

我们对工具变量估计的论述严格遵照我们在第 1 篇中对普通最小二乘的推导，其中假定我们有一个来自基本总体的随机样本。这个起点很合人意，因为除了简化符号之外，它还强调了应根据基本总体来表述对 IV 估计所做的重要的假定（正如用 OLS 时一样）。如我们在第 2 篇中所示，OLS 可以应用于时间序列数据，而工具变量法也一样可以。15.7 节讨论 IV 法应用于时间序列数据时出现的一些特殊问题。在 15.8 节中，我们将论述在混合横截面和综列数据上的应用。 462

15.1 动机：简单回归模型中的遗漏变量

面对可能发生的遗漏变量偏误（或未观测到的异质性），迄今为止我们已讨论了三种选择：（1）我们可以忽略此问题，承受有偏、非一致性估计量的后果；（2）我们可以试图为未观测到的变量寻找并使用一个适宜的代理变量；（3）我们可以假定遗漏变量不随时间变化，运用第 13 章与第 14 章中的固定效应或一阶差分方法。若能把估计值与关键参数的偏误方向一同给出，则第一个回答是令人满意的。例如，如果我们能说一个正参数（譬如职业培训对往后工资的影响）的估计量有朝零偏误，并且找到了一个统计上显著的正的估计值，那么我们还是学到了一些东西：职业培训对工资有正的影响，而我们很可能低估了该影响。不幸的是，相反的情况经常发生，我们的估计值可能在数值上太大了，以致我们要得出任何有用的结论都非常困难。

9.2 节中讨论的代理变量解也能获得令人满意的结果，但并不是总可以找到一个好的代理。该方法试图通过用代理变量取代不可观测的变量，来解决遗漏变量的问题。

另一种方法是将未观测到的变量留在误差项中，但不是用 OLS 估计模型，而是运用一种承认存在遗漏变量的估计方法。这便是工具变量法所要做的。

举例来说，考虑成年劳动者的工资方程中存在未观测到的能力的问题。一个简单的模型为

$$\log(wage)=\beta_0+\beta_1 educ+\beta_2 abil+e$$

式中，e 为误差项。在第 9 章中，我们说明了在某些假定下，如何用诸如 IQ 的代理变量代替能力，从而通过以下回归可得到一致性估计量：

$\log(wage)$ 对 $educ$，IQ 回归

然而，假定不能得到适当的代理变量（或它不具备足以获取一致性估计量所需的性质）。这样一来，我们将 $abil$ 放入误差项中，留下来的就是简单的回归模型：

$$\log(wage)=\beta_0+\beta_1 educ+u \tag{15.1}$$

式中，u 包含了 $abil$。当然，如果用 OLS 估计方程（15.1），若是 $educ$ 与 $abil$ 相关，得到的结果将是 β_1 的有偏、非一致性估计量。

最后证明是，假如能为 $educ$ 找到一个工具变量，我们仍可以根据方程（15.1）来进行估计。为描述该方法，将简单回归模型写成：

463

$$y=\beta_0+\beta_1 x+u \tag{15.2}$$

其中我们认为 x 与 u 相关：

$$\mathrm{Cov}(x,u)\neq 0 \tag{15.3}$$

工具变量法无论 x 与 u 相关与否都行得通，但是，如果 x 与 u 不相关，我们应该使用 OLS，其原因将在后面看到。

为了获得 x 与 u 相关时 β_0 和 β_1 的一致性估计量，我们还需要一些另外的信息。这些信息由一个满足某些性质的新变量给出。假定我们有一个可观测到的变量 z，它满足两个假定：(1) z 与 u 不相关，即

$$\mathrm{Cov}(z,u)=0 \tag{15.4}$$

(2) z 与 x 相关，即

$$\mathrm{Cov}(z,x)\neq 0 \tag{15.5}$$

我们则称 z 是 x 的**工具变量**（instrumental variable）。

有时候，人们把所需条件（15.4）概括为"z 在方程（15.2）中是外生的"。从遗漏变量的角度看，这意味着 z 应当对 y 无偏效应，也不应当与其他影响 y 的因素相关。方程（15.5）意味着 z 必然与内生解释变量 x 有着正的或负的关系。

对工具变量的两个要求之间有一个非常重要的差别。因为式（15.4）是 z 与不可观测的误差 u 的协方差，我们无法对它进行验证或哪怕是检验：我们必须求助于经济行为或内心感受来维持这一假定。相比之下，给定一个来自总体的随机样本，z 与 x 相关（在总体中）的条件则可加以检验。做到这一点最容易的方法是估计一个 x 与 z 之间的简单回归。在总体中，我们有

$$x=\pi_0+\pi_1 z+v \tag{15.6}$$

从而，由于 $\pi_1=\mathrm{Cov}(z,x)/\mathrm{Var}(z)$，式（15.5）中的假定当且仅当 $\pi_1\neq 0$ 时成立。因而我们就能够以充分小（习惯说充分高——译者注）的显著水平（5%或1%）拒绝虚拟假设

$$\mathrm{H}_0:\pi_1=0 \tag{15.7}$$

并接受双侧对立假设 H_0：$\pi_1\neq 0$。如果真是这样，我们能相当有把握肯定式（15.5）是成立的。

对于式（15.1）中的 log（$wage$）方程，$educ$ 的工具变量 z 必须：(1) 与能力（以及其他影响工资的不可观测的因素）不相关；(2) 与教育相关。诸如一个人的社会福利登记号的最后一位数字之类的变量，几乎一定满足第

一个必要条件：与能力不相关，因为它是随机决定的。然而，该变量与教育
464 不相关，因而是 *educ* 的一个低劣的工具变量。

我们所谓的用于遗漏变量的代理变量因相应的原因成为低劣的 IV。例如，在遗漏能力的 log（*wage*）例子中，*abil* 的代理变量应该尽可能地与 *abil* 高度相关，而工具变量必须与 *abil* 不相关。因此，尽管 *IQ* 是 *abil* 的一个好的代理变量候选者，它却不是 *educ* 的好的工具变量。

对其他可能的工具变量候选者，这些必要条件更加不确定。劳动经济学家已在工资方程中使用家庭背景变量作为教育的 IV。例如，母亲的教育（*motheduc*）与孩子的教育是正相关的，这一点通过搜集劳动者数据样本并做 *educ* 对 *motheduc* 的简单回归便可以看出来。因此，*motheduc* 满足方程（15.5）。问题是，母亲的教育也可能与孩子的能力相关（通过母亲的能力和可能通过孩子幼年所受的教养的质量）。

式（15.1）中 *educ* 的另一个 IV 选择是成长过程中兄弟姊妹的数目（*sibs*）。一般地说，较多的兄弟姊妹与较低的平均教育水平相联系。这样，如果兄弟姊妹的数目与能力不相关，它可以充当 *educ* 的工具变量。

再举一个例子，考虑估计逃课对期末考试成绩的因果影响的问题。在一个简单的回归框架中，我们有

$$score = \beta_0 + \beta_1 skipped + u \qquad (15.8)$$

式中，*score* 为期末考试成绩；*skipped* 为该学期逃课的总数目。当然，我们可能担心 *skipped* 与 u 中其他因素相关：较好的学生可能逃课较少。因而 *score* 对 *skipped* 的简单回归可能不会给我们一个对逃课的因果影响的好的估计。

什么可能是 *skipped* 的好的 IV？我们所需要的是对 *score* 无直接效应且与学生能力不相关的 IV。同时，该 IV 必须与 *skipped* 相关。一个选择是利用住宿区与学校之间的距离。一所大规模的大学中将有部分学生乘车去学校，这也许会增加逃课的可能性（由于恶劣的天气、睡过头等等）。因而，*skipped* 可能与 *distance* 正相关；这一点可通过 *skipped* 对 *distance* 的回归并做一个 t 检验得以验证，正如前面所描述的。

distance 是否与 u 不相关？在简单回归模型（15.8）中，u 中的一些因素可能与 *distance* 相关。例如，低收入家庭的学生可能不住在学校；如果收入影响到学生的行为，可能会导致 *distance* 与 u 相关。15.2 节说明如何在多元回归的情况下使用 IV，以便其他影响 *score* 的因素能直接包含在模型中。那么，*distance* 也许是 *skipped* 的一个好的 IV。如果学生能力有一个好的代理，例如以往学期的累积 *GPA*，IV 法可能根本就不需要。

现在我们来证明可得到的工具变量能够用于进行方程（15.2）中的一致
465 性参数估计。特别地，我们将说明式（15.4）与式（15.5）［等价地，式（15.4）与式（15.7）］中的假定足以识别参数 β_1。在这一点上，参数的**识别**（identification）意味着我们可以根据总体矩写出 β_1，总体矩可用样本数据来估计。为了根据总体协方差写出 β_1，我们利用方程（15.2）：x 与 y 之

间的协方差为

$$\mathrm{Cov}(z,y)=\beta_1\mathrm{Cov}(z,x)+\mathrm{Cov}(z,u)$$

现在，在式（15.4）中 $\mathrm{Cov}(z,u)=0$ 与式（15.5）中 $\mathrm{Cov}(z,x)\neq 0$ 的假定下，我们可以解出 β_1 为

$$\beta_1=\frac{\mathrm{Cov}(z,y)}{\mathrm{Cov}(z,x)} \tag{15.9}$$

[注意：如果 z 与 x 不相关，即 $\mathrm{Cov}(z,x)=0$，该简单代数式不成立。] 方程（15.9）表明 β_1 是 z，y 之间的总体协方差除以 z，x 之间的总体协方差的商，这说明了 β_1 被识别。给定一个随机样本，我们用对应样本量来估计总体的量。在分子和分母中约去样本容量后，我们得到 β_1 的**工具变量（IV）估计量** [instrumental variables (IV) estimator]：

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})}{\sum_{i=1}^{n}(z_i-\bar{z})(x_i-\bar{x})} \tag{15.10}$$

给定 x，y 和 z 的样本数据，很容易获得式（15.10）中的 IV 估计量。β_0 的 IV 估计量就为：$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$，除了其中的斜率估计量 $\hat{\beta}_1$ 现在为 IV 估计量，它看起来就像 OLS 中的截距估计量。

当 $z=x$ 时，我们获得 β_1 的 OLS 估计量绝不是偶然的。换句话说，当 x 是外生的时，它可用做自身的 IV，IV 估计量等同于 OLS 估计量。

大数定律的一个简单应用表明，如果满足式（15.4）和式（15.5）中的假定，β_1 的 IV 估计量具有一致性：$\mathrm{plim}(\hat{\beta}_1)=\beta_1$。如果任一个假定不成立，IV 估计量都将是非一致性的（这一点后面将进一步地研究）。IV 估计量的一个特点是：当事实上 x 与 u 相关——以致确实需要工具变量来估计——它实质上绝不是无偏的。在小样本中，这意味着 IV 估计量可能有相当大的偏误，这就是为什么希望有大样本的一个原因。

用 IV 估计量做统计推断

已知 IV 和 OLS 具有类似的结构，我们无须惊讶在大样本容量的情况下 IV 估计量近似服从正态分布。为了对 β_1 进行推断，我们需要一个可用于计算 t 统计量和置信区间的标准误，通常的方法是增加一个同方差性的假定，这和在 OLS 的情况下一样。不过现在，同方差性的假定是以工具变量 z，
466 而不是以内生解释变量 x 为条件来表述的。除了前面关于 u，x 和 z 的假定之外，我们增加

$$\mathrm{E}(u^2|z)=\sigma^2=\mathrm{Var}(u) \tag{15.11}$$

可以表明，在式（15.4）、式（15.5）和式（15.11）的假定下，$\hat{\beta}_1$ 的渐近方差为

$$\frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2} \tag{15.12}$$

式中，σ_x^2 为 x 的总体方差；σ^2 为 u 的总体方差；$\rho_{x,z}^2$ 为 x 与 z 之间的总体相关系数的平方，它告诉我们在总体中 x 与 z 是怎样地高度相关。如同运用 OLS 估计量一样，IV 估计量的渐近方差以 $\frac{1}{n}$ 的速度降为 0，这里 n 是样本容量。

方程（15.12）引起人们兴趣的原因有两点。第一，它提供了一种获得 IV 估计量的标准误的方法。方程（15.12）中的所有的量均可以在给定一个随机样本的情况下进行一致性的估计。为估计 σ_x^2，我们简单地计算出 x_i 的样本方差；为估计 $\rho_{x,z}^2$，我们可以做 x_i 对 z_i 的回归来获得 R^2，即 $R_{x,z}^2$。最后，为估计 σ^2，我们可以运用 IV 残差

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, 2, \cdots, n$$

式中，$\hat{\beta}_0$ 与 $\hat{\beta}_1$ 为 IV 估计量。σ^2 的一致性估计量看起来就像从简单 OLS 回归中得出的 σ^2 估计量：

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$$

其中用自由度进行纠正是标准的做法（即使随着样本容量的增加，这样做几乎不起什么作用）。

$\hat{\beta}_1$ 的（渐近的）标准误是所估计的渐近方差的平方根。这个渐近的方差由下式给出：

$$\frac{\hat{\sigma}^2}{\text{SST}_x \cdot R_{x,z}^2} \tag{15.13}$$

式中，SST_x 为 x_i 的总平方和。[回忆 x_i 的样本方差是 SST_x/n，因而约去样本容量我们得到（15.13）]。所得到的标准误可用于构造 t 统计量，以检验关于 β_1 的假设，或者是 β_1 的置信区间。$\hat{\beta}_0$ 也有一个标准误，我们在此不提。任何现代计量经济学的软件包都会计算出任一 IV 估计后的标准误。

在我们给出例子之前，比较 IV 和 OLS 估计量（当 x 与 u 不相关时）的渐近方差是有用处的。在高斯-马尔科夫假定下，OLS 估计量的方差为 σ^2/SST_x，而 IV 估计量类似的计算式为 $\sigma^2/\text{SST}_x \cdot R_{x,z}^2$；两者的区别仅在于 467 IV 的方差的分母中出现了 $R_{x,z}^2$。由于 R^2 总是小于 1，这个 2SLS 的方差总是大于 OLS 的方差（当 OLS 有效时）。如果 $R_{x,z}^2$ 很小，IV 的方差会比 OLS 的方差大得多。记住，$R_{x,z}^2$ 衡量的是样本中 x 与 z 之间的线性关系的大小。如果 x 与 z 只是轻度相关，$R_{x,z}^2$ 会很小，而这将转化为 IV 估计量的一个非常大的抽样方差。z 越是与 x 高度相关，$R_{x,z}^2$ 越是接近于 1，IV 估计量的方差就越小。在 $z = x$ 的情况下，$R_{x,z}^2 = 1$，我们得到 OLS 的方差，这正是所预期的。

前面的讨论突出了当 x 与 u 不相关时进行 IV 估计的一个重要代价：IV 估计量的渐近方差总是大于——有时大得很多——OLS 估计量的渐近方差。

例 15.1　对已婚女性进行教育的回报估计

我们用 MROZ.RAW 中关于已婚职业女性的数据来估计以下简单回归模型的教育回报：

$$\log(wage) = \beta_0 + \beta_1 educ + u \tag{15.14}$$

为了比较，我们首先得到 OLS 估计值：

$$\begin{aligned} \log(\widehat{wage}) &= -0.185 + 0.109 educ \\ &\quad (0.185) \quad (0.014) \\ &\quad n = 428, R^2 = 0.118 \end{aligned} \tag{15.15}$$

$\hat{\beta}_1$ 的估计值表明，再受一年的教育可得到约 11% 的回报。

接下来，我们用父亲的教育（*fatheduc*）作为 *educ* 的工具变量。我们必须认为 *fatheduc* 与 u 不相关。第二个必要条件是 *educ* 与 *fatheduc* 相关。做一个 *educ* 对 *fatheduc* 的简单回归（样本中只有职业女性），我们可以非常容易地验证这一点：

$$\begin{aligned} \widehat{educ} &= 10.24 + 0.269 fatheduc \\ &\quad (0.28) \quad (0.29) \\ &\quad n = 428, R^2 = 0.173 \end{aligned} \tag{15.16}$$

fatheduc 的 t 统计量为 9.28，说明 *educ* 与 *fatheduc* 之间存在统计上显著的正相关。（实际上，*fatheduc* 解释了样本中 *educ* 的变异中约 17% 的部分。）用 *fatheduc* 作为 *educ* 的 IV，得

$$\begin{aligned} \log(\widehat{wage}) &= -0.441 + 0.059 educ \\ &\quad (0.446) \quad (0.035) \\ &\quad n = 428, R^2 = 0.093 \end{aligned} \tag{15.17}$$

教育回报的 IV 估计值为 5.9%，大约是 OLS 估计值的 1/2。这表明 OLS 估计值过高，且与遗漏的能力变量的偏误相一致。但我们应该记住，这些都是 468 仅从一个样本中得出的估计值：我们根本不知道 0.109 是否高于真正的教育回报，或者 0.059 更接近真正的教育回报。其次，IV 估计量的标准误是 OLS 标准误的 2.5 倍（这是我们预期到的，原因已在前面表明）。运用 OLS 得出 β_1 的 95% 置信区间比运用 IV 要狭窄得多；事实上，IV 的置信区间确实包含了 OLS 估计值。因此，尽管实践中式（15.15）与式（15.17）之间的差异很大，我们不能说该差异在统计上显著。15.5 节中我们将说明如何对此进行检验。

在前面的例子中，运用 IV 估计出的教育回报小于运用 OLS 的估计结

果，这符合我们的预期。以下的例子将表明这个结果不是必然的。

例 15.2　估计对男性进行教育的回报

现在我们再用 WAGE2.RAW 来对男性估计教育的回报。我们用 *sibs*（兄弟姊妹的数目）作为 *educ* 的工具变量。它们是负相关的，对此我们可以从以下简单回归中来证实：

$$
\begin{aligned}
\widehat{educ} &= 14.14 - 0.228\,sibs \\
&\quad\ (0.11)\quad\ (0.030) \\
&\quad\ n = 935,\ R^2 = 0.057
\end{aligned}
$$

该方程意味着每多一个兄弟姐妹，相关联的是一年内所受的教育平均比原来减少约 0.23。如果我们假定 *sibs* 与式（15.14）中的误差项不相关，那么 IV 估计量就具有一致性。用 *sibs* 作为 *educ* 的工具变量估计方程（15.14），得

$$
\begin{aligned}
\log(\widehat{wage}) &= 5.13 + 0.122\,educ \\
&\quad\ (0.36)\quad (0.026) \\
&\quad\ n = 935
\end{aligned}
$$

（R^2 计算出为负数，因而我们没有予以报告。后面将从 IV 估计的角度对 R^2 进行讨论。）相比之下，β_1 的 OLS 估计值是 0.059，标准误是 0.006。与前面的例子不同，现在 IV 估计值比 OLS 估计值大得多。尽管我们不知道该差异是否在统计上显著，但它不会与 OLS 中遗漏的能力变量所造成的偏误相混淆（mesh with）。有可能 *sibs* 也与能力相关：较多的兄弟姊妹意味着平均起来受父母的照料较少，这可能导致较低的能力。另一个解释是，由于 *educ* 中的测量误差，OLS 估计量有朝零偏误。该解释不能完全令人信服，因为 *educ* 未必满足经典的含误差变量模型，这一点我们已在 9.3 节中讨论过。

在前面的例子中，内生解释变量（*educ*）与工具变量（*fatheduc*，*sibs*）均有数量含义。然而，这两类都可以是二值变量。安格里斯特和克鲁格（Angrist and Krueger，1991）在他们最简单的分析中，利用美国的男性人口 469 调查数据，提出了 *educ* 的一个巧妙的二值工具变量。如果该男性是在第一季度出生的，令 *frstqrt* 等于 1，否则为 0。式（15.14）中的误差项——特别是能力——似乎应该与出生季度不相关。但是，*frstqrt* 还要与 *educ* 相关。事实表明，在基于出生季度的总体中，教育年数确实有系统性差异。安格里斯特和克鲁格认为是缘于在各州实行的义务就学法，这很有说服力。简单地说，年初出生的学生往往入学较晚。因此，他们在达到义务教育年龄时（大部分州定为 16 岁），所受的教育略少于入学较早的学生。安格里斯特和克鲁格证实了，对于已完成高中学业的学生来说，受教育年数与出生季度并无关系。

因为教育年数在各出生季度之间的变化仅仅是微乎其微的——这意味着式（15.13）中的 $R^2_{x,z}$ 非常小——安格里斯特和克鲁格需要很大的样本容量

来得到一个合理而准确的 IV 估计值。利用 1920—1929 年之间出生的 247 199位男性的数据，得出教育回报的 OLS 估计值为0.080 1（标准误为 0.000 4），IV 估计值为0.071 5（0.021 9）；见于安格里斯特和克鲁格的论文中的表Ⅲ。注意到 OLS 估计值的 t 统计量那么大（约为 200），然而 IV 估计值的 t 统计量仅为 3.26。因而 IV 估计值在统计上不为零，但其置信区间比基于 OLS 估计值的置信区间宽得多。

安格里斯特和克鲁格有一个有趣的发现：IV 估计值与 OLS 估计值相差并不多。实际上，利用下一个十年中出生的男性的数据，得出 IV 估计值稍微高于 OLS 估计值。对此可以这样解释：说明在用 OLS 估计工资方程时不存在遗漏能力的偏误。可是，安格里斯特和克鲁格的论文在计量经济学界受到了非难。如同邦德、杰埃格和贝克（Bound, Jaeger and Baker, 1995）讨论的那样，它不能明显地判断出生季节与影响工资的诸因素不相关，纵然这些因素没有被人观测到。我们在下一小节中将解释，即使 z 与 u 之间有少量的相关，也会导致 IV 估计量存在严重的问题。

对于政策分析，内生解释变量往往是二值变量。例如，安格里斯特（Angrist, 1990）研究了参加越南战争的老兵其终身收入因参加越战而受到的影响。一个简单模型为

$$\log(earns) = \beta_0 + \beta_1 veteran + u \tag{15.18}$$

式中，*veteran* 为二值变量。疑问在于，用 OLS 估计该方程时，可能存在一个**自我选择**（self-selection）的问题，这一点我们在第 7 章中提到过：也许人们因为能从军队中得到最多的收入而选择参军，或者参军的决策与其他对收入有影响的特征相关。这些问题将导致 *veteran* 与 u 相关。

安格里斯特指出，越南战争的征兵抽签提供了一个**自然试验**（natural experiment）（亦参见第 13 章），从而产生了 *veteran* 的一个工具变量。年轻人被分给的征兵抽签号决定了他们是否会被征召去服役于越南战争。因为所分给的号码（毕竟）是随机分配的，征兵抽签号与误差项 u 不相关似乎是可信的。而得到号码足够小（指号码小于某个数——译者注）的人必须服役于越南战争，使得成为老兵的概率与抽签号相关。如果以上两点都是正确的，征兵抽签号是 *veteran* 的一个好的 IV 候选者。 470

问题 15.1 ☞ 如果某些被分给小的征兵抽签号的人，获得了更多的学校教育以减少了被征兵的概率，抽签号仍是式（15.18）中 *veteran* 的好的工具变量吗？

还有可能遇到一个二值的内生解释变量与一个二值的工具变量的情况。作为一个例子，参见习题 15.1。

低劣的工具变量条件下IV的性质

我们已经看到，尽管当 z 与 u 不相关，而 z 与 x 存在着正的或负的相关时，IV 是一致性的，但当 z 与 x 只是弱相关时 IV 估计值可能有大的标准误。z 与 x 之间的弱相关可能产生甚至是更加严重的后果：即使 z 与 u 只是适度相关，IV 估计量也会有大的渐近偏误。

当 z 与 u 可能相关时，通过对 IV 估计量的概率极限的分析，就可以看到这一点。利用总体相关和标准差，可以推出：

$$\text{plim}\hat{\beta}_1 = \beta_1 + \frac{\text{Corr}(z,u)}{\text{Corr}(z,x)} \cdot \frac{\sigma_u}{\sigma_x} \tag{15.19}$$

式中，σ_u 和 σ_x 分别代表总体中 u 和 x 的标准差。该方程中引起人们兴趣的是包含相关项的部分。它表明，即使 Corr(z,x)很小，如果 Corr(z,u)也很小，IV 估计量的非一致性会非常大。因此，即使我们只考虑一致性，如果 z 与 u 之间的相关小于 x 与 u 之间的相关，使用 IV 不一定比 OLS 更好。由于 Corr(x,u) = Cov$(x,u)/(\sigma_x\sigma_u)$，连同方程（5.3）一起，我们可以将 OLS 估计量的 plim——称之为 $\tilde{\beta}_1$——写为

$$\text{plim}\tilde{\beta}_1 = \beta_1 + \text{Corr}(x,u) \cdot \frac{\sigma_u}{\sigma_x} \tag{15.20}$$

比较两式，说明当时 Corr(z,u)/Corr(z,x) < Corr(x,u)，IV 就渐近偏误而言比 OLS 更可取。

在前面提到的安格里斯特和克鲁格（Angrist and Krueger，1991）的例子中，x 是学校教育的年数，z 是一个指示出生季度的二值变量，z 与 x 之间的相关非常小。邦德、杰埃格和贝克（Bound，Jaeger and Baker，1995）讨论了出生季度与 u 可能有些相关的原因。从方程（15.19）中，我们看到这将会导致 IV 估计量有相当大的偏误。

当 z 与 x 完全不相关时，无论 z 是否与 u 不相关，事情尤其糟糕。接下来的例子说明了为什么我们应当时常检查内生解释变量是否与备选的 IV 相关。

---

471 **例 15.3　估计吸烟对出生体重的影响**

在第 6 章中，我们估计了吸烟对婴儿出生体重的影响。没有其他的解释变量，模型为

$$\log(bwght) = \beta_0 + \beta_1 packs + u \tag{15.21}$$

式中，*packs* 为母亲每天吸烟的包数。我们会担心 *packs* 与其他健康因素或者获得良好的产前护理的可能性相关，以致 *packs* 与 u 可能相关。*packs* 的一个可能的工具变量是所居住州的香烟价格 *cigprice*。我们将假定 *cigprice* 与

u 不相关（即使州政府对健康护理的支持可能与香烟税相关）。

如果香烟是典型的消费品，基本的经济理论表明 *packs* 与 *cigprice* 负相关，所以 *cigprice* 可用做 *packs* 的 IV。为验证这一点，我们利用 BWGHT.RAW 中的数据，做 *packs* 对 *cigprice* 的回归：

$$\widehat{packs} = 0.067 + 0.000\,3 cigprice$$
$$(0.103)(0.000\,8)$$
$$n = 1\,338, R^2 = 0.000\,0, R^2 = -0.000\,6$$

这说明怀孕期间吸烟与香烟价格之间没有关系。考虑到吸烟有使人上瘾的特性，该结论可能不会太令人惊讶。

因为 *packs* 与 *cigprice* 不相关，我们不应该在式（15.21）中用 *cigprice* 作为 *packs* 的 IV，但如果我们用了会怎么样？IV 的结果将为

$$\log(\widehat{bwght}) = 4.45 + 2.99 packs$$
$$(0.91)(8.70)$$
$$n = 1\,388$$

（所报告的 R^2 为负数。）*packs* 的系数极大，而且有一个意想不到的符号。标准误也非常大，因此 *packs* 不是显著的。可是估计值是没有意义的，因为 *cigprice* 不满足我们总可以检验的 IV 的一个必要条件，即式（15.5）中的假定。

IV 估计后计算 R^2

大多数回归软件包运用标准公式 $R^2 = 1 - \mathrm{SSR/SST}$ 计算 IV 估计之后的 R^2，其中 SSR 是 IV 残差的平方和，SST 是 y 的总平方和。与 OLS 中的情况不同，由于 IV 的 SSR 实际上可能大于 SST，IV 估计中 R^2 的可能为负数，尽管报告 IV 估计的 R^2 不会有什么害处，但也不很有用。当 x 与 u 相关时，我们不能将 y 的方差分解成 $\beta_1^2\mathrm{Var}(x) + \mathrm{Var}(u)$，因此对 R^2 没有合理的解释。另外，正如我们将在 15.3 节中讨论的，这些 R^2 不能以通常的方法用于计算联合约束的 F 检验值。 472

如果我们的目标是要得出最大的 R^2，我们将总是用 OLS。IV 法是打算当 x 与 u 相关时，为 x 在其余条件不变情况下对 y 的影响提供更好的估计值；拟合优度不是考虑的因素。如果我们不能对 β_1 进行一致性估计，从 OLS 中得出高的 R^2 也不会让人感到欣慰。

15.2 多元回归模型的 IV 估计

简单回归模型的 IV 估计量容易延伸至多元回归的情形。我们从仅有一个解释变量与误差相关的情形开始。实际上，考虑两个解释变量条件下的标准线性模型：

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1 \tag{15.22}$$

我们称之为**结构方程**（structural equation），以强调我们的兴趣在于 β_j，这仅仅意味着此方程应该测量一个因果关系。在此我们用一个新的符号来区分内生变量与**外生变量**（exogenous variables）。因变量 y_1 显然是内生的，它与 u_1 相关。变量 y_2 和 z_1 是解释变量，u_1 是误差。通常，我们假定 u_1 的期望值为零：E（u_1）＝0。我们用 z_1 表示该变量在式（15.22）中是外生的（z_1 与 u_1 不相关）。我们用 y_2 表示该变量被怀疑与 u_1 相关。我们没有详细地说明为什么 y_2 与 u_1 相关，但现在最好认为 u_1 包含一个与 y_2 相关的遗漏变量。方程（15.22）中的符号源自于联立方程模型（我们将在第 16 章中讨论），但我们把它更广泛地用于多元回归模型中，目的是容易区分外生变量和内生变量。

式（15.22）的一个例子是

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u_1 \tag{15.23}$$

式中，$y_1 = \log$（$wage$），$y_2 = educ$，$z_1 = exper$。换句话说，我们假定 $exper$ 在式（15.23）中是外生的，但我们允许 $educ$——由于通常的原因——与 u_1 相关。

我们知道，如果用 OLS 估计式（15.22），所有的估计量将是有偏而非一致性的。这样，我们采用前一节中建议的策略，寻找 y_2 的工具变量。因为假定了 z_1 与 u_1 不相关，我们能否假定 y_2 与 z_1 相关而将 z_1 用做 y_2 的工具呢？答案是不能。既然 z_1 自身作为解释变量出现在式（15.22）中，它就不能用做 y_2 的工具变量。我们需要另外一个外生变量——称之为 z_2——它不出现在式（15.22）中。因此，关键的假定是 z_1，z_2 与 u_1 不相关；我们还假定 u_1 具有零均值，当方程包含截距时，这并不失普遍性。

478

$$\mathrm{E}(u_1) = 0, \mathrm{Cov}(z_1, u_1) = 0, \text{和 } \mathrm{Cov}(z_2, u_1) = 0 \tag{15.24}$$

给定零均值的假定，后两个假定等价于 $\mathrm{E}(z_1 u_1) = \mathrm{E}(z_2 u_1) = 0$，因而按照矩法的意思是求解式（15.24）的对应样本方程来获得 $\hat{\beta}_0$，$\hat{\beta}_1$ 和 $\hat{\beta}_2$：

$$\sum_{i=1}^{n}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) = 0$$

$$\sum_{i=1}^{n} z_{i1}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) = 0 \tag{15.25}$$

$$\sum_{i=1}^{n} z_{i2}(y_{i1}-\hat{\beta}_0-\hat{\beta}_1 y_{i2}-\hat{\beta}_2 z_{i1})=0$$

这是关于三个未知量 $\hat{\beta}_0$，$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的三线性方程组，给定 y_1，y_2，z_1 和 z_2 的数据，它很易于求解。这些估计量叫做工具变量估计量。如果我们认为 y_2 是外生的，并选择 $z_2=y_2$，方程（15.25）恰恰是OLS估计量的一阶条件；参见方程（3.13）。

我们仍需要工具变量 z_2 与 y_2 相关，可是这两个变量必须相关的含义因方程（15.22）中存在 z_1 而变得复杂。我们现在需要从偏相关的角度来表述这一假定。表述该条件最容易的方法是将内生解释变量写成关于外生变量和误差项的一个线性函数：

$$y_2=\pi_0+\pi_1 z_1+\pi_2 z_2+v_2 \tag{15.26}$$

其中，定义

$$\mathrm{E}(v_2)=0,\mathrm{Cov}(z_1,v_2)=0,\mathrm{Cov}(z_2,v_2)=0$$

π_j 为未知参数。关键的识别条件［除了方程（15.24）之外］是

$$\pi_2\neq 0 \tag{15.27}$$

换句话说，排除了 z_1 的影响后，y_2 与 z_2 仍然相关。该相关可正可负，但不为零。检验方程（15.27）是容易的：我们通过OLS估计方程（15.26），并运用 t 检验（也许要把它变换成即使出现异方差也能适用的、所谓对异方差性强健的 t 检验）。我们应当时常检验这一假定。不幸的是，我们不能检验 z_1 和 z_2 与 u_1 不相关；这一点必须不加怀疑地接受。

问题15.2 ☞

假定我们想要估计吸食大麻对大学平均积分点的影响。对于大学四年级学生构成的总体，令 *daysused* 表示过去的一个月中一个学生吸食大麻的天数，考虑结构方程

$$colGPA=\beta_0+\beta_1 daysused+\beta_2 SAT+u$$

（i）令 *percHS* 表示该学生的高中的同届毕业班中被报道定期吸食大麻的人数百分比。如果这是 *daysused* 的一个IV备选变量，写出 *daysused* 的诱导型。你认为式（15.27）可能是正确的吗？

（ii）你认为结构方程中的 *percHS* 真的是外生的吗？这里可能存在什么问题？

474 方程（15.26）是**诱导型方程**（reduced form equation）的一个例子，它意味着我们是用外生变量来表述内生变量的。这个名称源自于联立方程模型——我们将在下一章中进行研究——但是每逢有内生解释变量，它都是一个有用的概念，帮助我们把它和结构方程（15.22）区分开来。

在模型中增添更多的外生解释变量是简单易行的。将结构模型写成

$$y_1=\beta_0+\beta_1 y_2+\beta_2 z_1+\cdots+\beta_k z_{k-1}+u_1 \tag{15.28}$$

式中，y_2 被认为与 u_1 相关。令 z_k 也是一个外生变量，但它不在式 (15.28) 中。因此，我们假定

$$\mathrm{E}(u_1)=0, \mathrm{Cov}(z_j, u_1)=0, \quad j=1,\cdots,k \tag{15.29}$$

y_2 的诱导型是

$$y_2=\pi_0+\pi_1 z_1+\cdots+\pi_{k-1} z_{k-1}+\pi_k z_k+v_2 \tag{15.30}$$

我们需要 z_k 与 y_2 之间存在某些偏相关：

$$\pi_k \neq 0 \tag{15.31}$$

在式 (15.29) 和式 (15.31) 中的假定下，z_k 是 y_2 的一个有效的 IV。（我们不关心其余的 π_j；它们可能部分或全部为零。）考虑 z_1，…，z_{k-1} 可用做它们自身的 IV 是合理的；因此，外生变量系列常常叫做工具变量系列。一个次要的补充假定是，外生变量之间不存在完全线性关系；这类似于 OLS 情况下的非完全共线性假定。

对于标准的统计推断，我们需要假定 u_1 具有同方差性。15.3 节中我们将在更一般的环境下仔细地表述这些假定。

例 15.4　用邻近大学作为教育的 IV

卡德（Card，1995）利用 1976 年的工资和教育的一个男性样本数据来估计教育回报。他运用这样一个虚拟变量作为教育的工具变量，即是否在一所四年制大学的附近长大（*nearc4*）。在一个 log（*wage*）方程中，还引进了其他标准控制变量：经验、黑人的虚拟变量、居住在大城市及其郊区（SMSA）和居住在南方的虚拟变量，一整套地域性的虚拟变量以及 1966 年在何处居住的 SMSA 虚拟变量。为了 *nearc4* 成为一个有效的工具，它必须与工资方程中的误差项不相关——我们假定如此——且必须与 *educ* 偏相关。为验证后一个所需条件，我们将 *educ* 对 *nearc4* 及方程中出现的所有外生变量做回归。（那就是说，我们估计 educ 的诱导型。）利用 CARD.RAW 中的数据，以缩写形式我们获得

475

$$\begin{aligned}\widehat{educ}&=16.64+0.320\,nearc4-0.413\,exper+\cdots\\&\quad\;(0.24)\;\;(0.088)\qquad\;\;(0.034)\\n&=3\,010, R^2=0.477\end{aligned} \tag{15.32}$$

我们的兴趣在于 *nearc4* 的系数及其 t 统计量。其系数意味着，在其他因素（经历、种族、地域等）固定的情况下，曾于 1966 年住在大学附近的人 1976 年所受的教育比不在大学附近长大的人平均多出约 1/3。*nearc4* 的 t 统计量是 3.64，其对应的 p 值在小数点后的前三位数字均为零。因此，如果 *nearc4* 与误差项中未观测到的因素不相关，我们就可以用 *nearc4* 作为 *educ* 的 IV。

OLS 和 IV 估计值由表 15.1 给出。有趣的是，教育回报的 IV 估计值将近是 OLS 估计值的 2 倍，而 IV 估计值的标准误却比 OLS 的标准误大 18 倍

还多。IV 估计值的 95% 置信区间是从 0.024～0.239，这是一个很宽的范围。当我们认为 *educ* 是内生的时，要得到教育回报的一致性估计量所必须付出的代价将是更大的置信区间。

476

表 15.1 因变量：log(*wage*)

解释变量	OLS	IV
educ	0.075 (0.003)	0.132 (0.055)
exper	0.085 (0.007)	0.108 (0.024)
$exper^2$	−0.002 3 (0.000 3)	−0.002 3 (0.000 3)
black	−0.199 (0.018)	−0.147 (0.054)
smsa	0.136 (0.020)	0.112 (0.032)
south	−0.148 (0.026)	0.145 (0.027)
观测次数	3 010	3 010
R^2	0.300	0.238

其他控制变量：*smsa*66，*reg*662，…，*reg*669

正如前面讨论的，我们会认为在 IV 估计中，较小的 R^2 并不奇怪：按照定义，由于 OLS 使残差平方和最小化，OLS 的 R^2 将总是大一些。

15.3 两阶段最小二乘

在前一节中，我们假定有单一的内生解释变量 y_2 和 y_2 的一个工具变量。可往往我们有不止一个外生变量，它们被排斥在结构模型之外，且可能与 y_2 相关，这意味着它们是 y_2 的有效的 IV。在本节中，我们讨论如何运用复工具变量。

单一内生解释变量

重新考虑结构模型（15.22），它有一个内生解释变量和一个外生解释变

量。假定现在我们有两个被排斥在式（15.22）之外的外生变量：z_2 和 z_3。z_2 和 z_3 不出现在式（15.22）中，且与误差项不相关的诸假定称为**排斥性约束**（exclusion restrictions）。

如果 z_2 和 z_3 都与 y_2 相关，我们就可仅用任一个变量作为 IV，如同前一节那样。但这样一来，我们将有两个 IV 估计量，而一般地说没有一个会是有效的。由于 z_1，z_2 和 z_3 各自与 u_1 不相关，它们的任何线性组合也与 u_1 不相关，因此，外生变量的任何线性组合都是有效的 IV。为寻找最好的 IV，我们选择与 y_2 最高度相关的线性组合。这正是由 y_2 的诱导型方程所给出的。写

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2 \tag{15.33}$$

其中

$$\mathrm{E}(v_2) = 0, \mathrm{Cov}(z_1, v_2) = 0, \mathrm{Cov}(z_2, v_2) = 0, \mathrm{Cov}(z_3, v_2) = 0$$

那么，y_2 最好的 IV（在本章附录中给出的假定下）是式（15.33）中 z_j 的线性组合，我们称之为 y_2^*：

$$y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 \tag{15.34}$$

为了使该 IV 与 z_1 不是完全相关，我们需要 π_2 或 π_3 之中至少一个不为零：

$$\pi_2 \neq 0 \text{ 或 } \pi_3 \neq 0 \tag{15.35}$$

一旦我们假定 z_j 全部都是外生的，这便是关键的识别假定。（π_1 的值是不
477 相干的。）如果 $\pi_2 = 0$ 且 $\pi_3 = 0$，结构方程（15.22）将不被识别。我们可以运用 F 统计量，检验 H_0：$\pi_2 = 0$ 与 $\pi_3 = 0$，其对立假设为式（15.35）。

以一个有用的方式来考虑式（15.33），将 y_2 分成两部分。第一部分是 y_2^*，这是 y_2 中与误差项 u_1 不相关的部分。第二部分是 v_2，它可能与 u_1 相关——这是 y_2 可能内生的原因。

已知 z_j 的数据，假如我们知道总体参数 π_j，我们可对每次观测计算 y_2^*。在实践中这根本不真实。然而，正如在前一节中所看到的，我们总是可以用 OLS 估计诱导型。这样，利用样本，我们将 y_2 对 z_1，z_2 和 z_3 回归，获得拟合值：

$$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3 \tag{15.36}$$

（就是说，对每个 i，我们有 $\hat{y}_{i2}$）。现在，我们将证实在式（15.33）中 z_2 与 z_3 以一个相当小的显著水平（不大于 5%）联合显著。如果 z_2 与 z_3 在式（15.33）中不是联合显著的，做 IV 估计是在浪费时间。

一旦我们有了 $\hat{y}_2$，我们便可以用它作为 y_2 的 IV。用于估计 β_0，β_1 和 β_2 的三个方程是（15.25）中的前两个方程和代替第三个方程的

$$\sum_{i=1}^{n} \hat{y}_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) = 0 \tag{15.37}$$

求解关于三个未知量的三个方程，我们得到 IV 估计量。

在复工具条件下，IV 估计量也叫做**两阶段最小二乘（2SLS）估计量**［two stage least squares (2SLS) estimator］。原因很简单。运用 OLS 代数，可以说明当我们用 $\hat{y}_2$ 作为 y_2 的 IV 时，IV 估计值 $\hat{\beta}_0$，$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 等同于从

$$y_1 \text{ 对 } \hat{y}_2 \text{ 和 } z_1 \tag{15.38}$$

的回归中得出的 OLS 估计值。换句话说，我们可以通过两阶段来获得 2SLS。第一阶段是做方程（15.36）中的回归，我们得到拟合值 $\hat{y}_2$。第二阶段是做方程（15.38）中的 OLS 回归。因为我们用 $\hat{y}_2$ 代替了 y_2，2SLS 估计值与 OLS 估计值有实质上的差异。

一些经济学家喜欢这样来解释式（15.38）中的回归：拟合值 $\hat{y}_2$ 是 y_2^* 的估计形式，y_2^* 与 u_1 不相关。因此，2SLS 在做式（15.38）的 OLS 回归之前先“清除”y_2 中与 u_1 的相关。这一说法，可通过将 $y_2 = y_2^* + v_2$ 代入式（15.22）中，发现其正确性：

$$y_1 = \beta_0 + \beta_1 y_2^* + \beta_2 z_1 + u_1 + \beta_1 v_2 \tag{15.39}$$

现在合成误差 $u_1 + \beta_1 v_2$ 有零均值，且与 y_2^* 和 z_1 不相关，这就是 OLS 估计有效的原因。

大多数计量经济学的软件包对 2SLS 有专门的指令，所以无须明确地分两阶段进行。实际上，在大多数情况下，你应当避免用手工来做第二阶段的
478 工作，因为以这样的方法获得的标准误和检验统计量是不正确的。［原因是式（15.39）中的误差项包括 v_2，但标准误只包括 u_1 的方差。］任何支持 2SLS 的回归软件要求有因变量、解释变量（内生和外生）和整个工具变量的表列（即所有的外生变量），其结果典型地说十分类似于 OLS 的结果。

在 y_2 有单一 IV 的模型（15.28）中，15.2 节得出的 IV 估计量等同于 2SLS 估计量。因此，当我们对每个内生解释变量有一个 IV 时，我们可称估计方法为 IV 或 2SLS。

增添更多外生变量，方法上也没有什么变化。例如，假定工资方程为

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_1 \tag{15.40}$$

式中，u_1 与 $exper$ 和 $exper^2$ 均不相关。假定我们还认为母亲和父亲的教育与 u_1 不相关，那么可以将它们都用做 $educ$ 的 IV。$educ$ 的诱导型方程为

$$educ = \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 motheduc + \pi_4 fatheduc + v_2 \tag{15.41}$$

识别的要求是 $\pi_3 \neq 0$ 或 $\pi_4 \neq 0$（或两个都非零）。

例 15.5 职业女性的教育回报

用 MROZ.RAW 中的数据估计方程（15.40）。首先，我们在式（15.41）中用 F 检验来检验 H_0：$\pi_3 = 0$，$\pi_4 = 0$。结果是 $F = 55.40$，p 值 $= 0.000\ 0$。

正如所预期的，*educ* 与父母的教育（偏）相关。

当我们用 2SLS 估计式（15.40）时，获得的方程形式为

$$\begin{aligned}\log(\widehat{wage}) &= \underset{(0.400)}{0.048} + \underset{(0.031)}{0.061}\,educ + \underset{(0.013)}{0.044}\,exper - \underset{(0.0004)}{0.0009}\,exper^2\\ n &= 428,\ R^2 = 0.136\end{aligned}$$

所估计的教育回报约为 6.1%，相比 OLS 估计值约为 10.8%。由于它相对大的标准误，在对应着双侧对立假设的 5% 的显著水平上，2SLS 估计值几乎不显著。

本章附录中给出了 2SLS 需要的假定，它们使得 2SLS 具备所希望的大样本性质。但在此进行简要的概述是有用的。如果我们写出如式（15.28）中的结构方程

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1 \tag{15.42}$$

然后我们假定每个 z_j 与 u_1 不相关。另外我们至少需要一个与 y_2 偏相关的外生变量不在式（15.42）之中。这保证了一致性。为了使通常的 2SLS 标准误 479 和 t 统计量渐近有效，我们还需要一个同方差性的假定：结构误差 u_1 的方差不会与任何外生变量有关。对于时间序列上的应用，我们需要更多的假定，这将在 15.7 节中看到。

多重共线性与2SLS

在第 3 章中，我们介绍了多重共线性的问题，并说明回归元之间的相关如何导致 OLS 估计值具有大的标准误。多重共线性在 2SLS 条件下甚至会更严重。要知道为什么，可以把 β_1 的 2SLS 估计量的（渐近）方差近似地写为

$$\frac{\sigma^2}{\widehat{\mathrm{SST}}_2(1-\hat{R}_2^2)} \tag{15.43}$$

式中，$\sigma^2 = \mathrm{Var}(u_1)$；$\widehat{\mathrm{SST}}_2$ 为 $\hat{y}_2$ 中的总变异；$\hat{R}_2^2$ 为将 $\hat{y}_2$ 对其他所有出现在结构方程中的外生变量做回归得出的 R^2。2SLS 的方差大于 OLS 的方差的原因有两点。第一，$\hat{y}_2$ 从结构上看，其变异比 y_2 小。（记住，总平方和 = 解释平方和 + 残差平方和；y_2 中的变异构成总平方和，而 $\hat{y}_2$ 中的变异构成解释平方和。）第二，$\hat{y}_2$ 与式（15.42）中外生变量之间的相关往往比 y_2 与这些变量之间的相关大得多。这在本质上解释了 2SLS 中的多重共线性问题。

作为一个实例来考虑例 15.4。当 *educ* 对表 15.1 中的外生变量做回归时，$R^2 = 0.475$；这是中等程度的多重共线性，但重要的是 $\hat{\beta}_{educ}$ 的 OLS 标准误差相当小。当我们获得第一阶段的拟合值 $\widehat{educ}$，并将它们对表 15.1 中外生变量做回归时，$R^2 = 0.995$，这表明 $\widehat{educ}$ 与表中其余的外生变量之间有

很高程度的多重共线性。(这个高的 R^2 并不太令人吃惊，因为 *educ* 是关于表 15.1 中所有外生变量和 *nearc*4 的一个函数。）方程（15.43）表明，接近 1 的 R_2^2 可导致 2SLS 估计值有非常大的标准误。然而如在 OLS 条件下一样，大样本容量可帮助抵消大的 R_2^2。

多个内生解释变量

两阶段最小二乘也可以用于不止一个内生解释变量情形下的模型中。例如，考虑模型

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + u_1 \qquad (15.44)$$

式中，$E(u_1)=0$；u_1 与 z_1，z_2 和 z_3 不相关。变量 y_2 和 y_3 是内生解释变量：每个都可能与 u_1 相关。

为了用 2SLS 估计式（15.44），我们需要至少两个外生变量，它们不出现在式（15.44）中，但与 y_2 和 y_3 相关。假定我们有两个被排斥的外生变
480 量，即 z_4 和 z_5。然后，根据我们对单一内生解释变量的分析，我们需要 z_4 或者 z_5 出现在 y_2 和 y_3 的诱导型中。(与前面一样，我们可以用 F 统计量来检验。）尽管这对于识别是必要的，但不幸的是，它不是充分的。假定 z_4 出现在每个诱导型中，而 z_5 在两个中都没有出现。那么，我们并不是真正的有两个外生变量与 y_2 和 y_3 偏相关。两阶段最小二乘不会获取 β_j 的一致性估计量。

一般来说，当我们在回归模型中有不止一个内生解释变量时，在若干复杂的情况下仍可能不能识别。但是，我们可以容易地表述识别的一个必要条件，叫做**阶条件**（order condition）。

方程识别的阶条件：我们需要被排斥的外生变量至少与结构方程中包括的内生解释变量一样多。验证阶条件是简单的，因为它只须数一数内生和外生变量的个数。识别的充分条件称为**秩条件**（rank condition）。我们在前面已见到不少秩条件的特例——例如，围绕方程（15.35）的讨论。对秩条件的一般表述需要矩阵代数，超出了本书的范围。［参见 Wooldridge（1999，第 5 章）。］

问题 15.3☞ 以下模型用一个是否存在枪支管理法的二值变量及其他控制变量，以解释城市的暴力犯罪率：

$$\begin{aligned} violent &= \beta_0 + \beta_1 guncontrol + \beta_2 unem + \beta_3 popul \\ &\quad + \beta_4 percblck + \beta_5 age18\text{–}21 + \cdots \end{aligned}$$

一些研究者估计了类似的方程，他们运用诸如国家步枪协会中城市会员的数目、枪支杂志订阅者的数目等变量作为 *guncontrol* 的工具变量。［例如，参见 Kleck 和 Patterson（1993)］。它们是令人信服的工具吗？

2SLS 估计后对多个假设的检验

在一个用 2SLS 来估计的模型中，检验多个假设时我们必须小心。正如我们在第 4 章中 OLS 条件下所学过的，运用残差平方和或 F 统计量的 R^2 形式是很吸引人的。然而，2SLS 中的 R^2 可能为负数的事实表明，通常计算 F 统计量的方法可能不适合；现在就遇到了这个问题。实际上，如果我们用 2SLS 残差去计算受约束和无约束模型的 SSR，不能保证 $\mathrm{SSR}_r \geqslant \mathrm{SSR}_{ur}$；如果反之成立，$F$ 估计量将为负数。

有可能将第二阶段回归［例如式（15.38）］得出的残差平方和与 SSR_{ur} 结合起来，以获得一个在大样本下近似服从 F 分布的统计量。因为许多计量经济学软件包中有使用简单的检验指令，它们可用于检验 2SLS 估计后的多个假设，这里不作详细介绍。戴维森和麦金农（Davidson and Mackinnon, 1993）和伍德里奇（Wooldridge, 1999，第 5 章）含有如何计算 2SLS 的 F 统计量的讨论。

15.4 含误差的变量问题的 IV 解

在前一节中，我们提出用工具变量作为解决遗漏变量问题的方法，然而它们也能用于处理测量误差的问题。为了说明，考虑模型

481

$$y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u \qquad (15.45)$$

式中，y 和 x_2 是可观测到的，而 x_1^* 则观测不到。令 x_1 是 x_1^* 的一个可观测到的度量：$x_1 = x_1^* + e_1$（e_1 是测量误差）。在第 9 章中，我们说明了 x_1 与 e_1 之间的相关导致了 OLS 的有偏和非一致性，这里用 x_1 代替了 x_1^*。写出下式，就可以看到这一点

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (u - \beta_1 e_1) \qquad (15.46)$$

如果经典的含误差变量（CEV）的假定成立，β_1 的 OLS 估计量有朝零偏误。没有进一步的假定，我们对此毫无办法。

在一些情况下，我们可以用 IV 方法来解决测量误差问题。在式（15.46）中，我们假定 u 与 x_1^*、x_1 和 x_2 不相关；在 CEV 的情况下，我们假定 e_1 与 x_1^* 和 x_2 不相关。这些意味着 x_2 在式（15.46）中是外生的，可是 x_1 与 e_1 相关。我们所需的是 x_1 的 IV。这样的 IV 必须与 x_1 相关，与 u 不相关——从而它必须被排斥在式（15.45）之外——并且与测量误差 e_1 不相关。

一种可能是获取 x_1^* 的第二个度量，即 z_1。既然影响 y 的是 x_1^*，假定 z_1 与 u 不相关是自然不过的了。如果我们写成 $z_1 = x_1^* + a_1$（a_1 是 z_1 的测量误差），那么我们必须假定 a_1 与 e_1 不相关。换句话说，x_1 和 z_1 都错误地测量了 x_1^*，但它们的测量误差不相关。当然，x_1 和 z_1 通过对 x_1^* 的相依而相关，因而我们可以用 z_1 作为 x_1 的 IV。

什么时候我们可以得到一个变量的两个度量呢？有时，当一群工人被问及他们的年薪时，他们的雇主可以提供第二个度量。对于夫妻俩，每一方都可以独立地报告储蓄或家庭收入的水平。14.3 节所引用的阿申费尔特和克鲁格（Ashenfelter and Krueger，1994）的研究中，每个双胞胎被问及他的兄弟或她的姊妹所受教育的年数；这给出了第二个度量，它可以在工资方程中用做自我报告的教育的 IV。（阿申费尔特和克鲁格还结合差分和 IV 来解释遗漏变量问题；对此 15.8 节中有更多的讨论。）然而一般地说，一个解释变量有两个度量是罕见的。

另一个选择是运用其他外生变量，将它们作为潜在的误测变量的 IV。例如，我们在例 15.5 中用 *motheduc* 和 *fatheduc* 作为 *educ* 的 IV，可以达到该目的。如果我们认为 $educ = educ^* + e_1$，若 *motheduc* 和 *fatheduc* 与测量误差 e_1 不相关，那么例 15.5 中的 IV 估计值不会受测量误差的影响。比起假定 *motheduc* 和 *fatheduc* 与能力不相关，而能力却包含在式（15.45）的 u 中，这可能更加合理。

当运用像测验成绩等去控制未观测到的特征时，也可以用 IV 方法。在 9.2 节中，我们说明了，在某些假定下，代理变量可用于解决遗漏变量问题。482 例 9.3 中，我们用 *IQ* 作为未观测到的能力的代理变量。这仅仅需要在模型中添加 *IQ* 并做一个 OLS 回归。但是当 *IQ* 不完全满足代理变量的假定时，存在另一种行之有效的选择。举例说明，将工资方程写成

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + abil + u \qquad (15.47)$$

这里我们又一次看到遗漏变量的问题。可是我们有两种测验成绩作为能力的指标。我们假定成绩可写为

$$test_1 = \gamma_1 abil + e_1$$

和

$$test_2 = \delta_1 abil + e_2$$

式中，$\gamma_1 > 0$；$\delta_1 > 0$。既然影响工资的是能力，我们就可以假定 $test_1$ 和 $test_2$ 与 u 不相关。如果我们根据第一种测验成绩写出 *abil*，并将之代入式（15.47），得到

$$\begin{aligned}\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 \\ + \alpha_1 test_1 + (u - \alpha_1 e_1)\end{aligned} \qquad (15.48)$$

式中，$\alpha_1 = 1/\gamma_1$。现在，如果我们假定 e_1 与式（15.47）中包括 *abil* 在内的所有解释变量不相关，那么 e_1 与 $test_1$ 必须相关。［注意到 *educ* 在式

(15.47) 中不是内生的；而 $test_1$ 是。] 这意味着用 OLS 估计式 (15.48) 将得到 β_j（和 α_1）的非一致性估计量。在我们所做的假定下，$test_1$ 不满足代理变量的假定。

如果我们假定 e_2 也与式 (15.47) 中的所有解释变量不相关，并且 e_1 与 e_2 不相关，那么 e_1 与第二种测验成绩 $test_2$ 不相关。因此，$test_2$ 可用做 $test_1$ 的 IV。

例 15.6　用两种测验成绩作为能力的指标

我们利用 WAGE2. RAW 中的数据实施前面的程序，其中 *IQ* 起着第一种测验成绩的作用，*KWW*（工作领域中的知识）是第二种测验成绩。解释变量与例 9.3 中的一样：*educ*，*exper*，*tenure*，*married*，*south*，*urban* 和 *black*。我们不是像表 9.2 第 (2) 列中那样添加 *IQ* 做 OLS，而是添加 *IQ*，并用 *KWW* 作为它的工具。*educ* 的系数是 0.025 (se = 0.017)。这是个低的估计值，且在统计上无异于零。该发现是令人费解的，它表明我们的诸假定之一是不成立的；也许 e_1 与 e_2 相关。

15.5　内生性检验与检验过度识别约束

在本节中，我们根据工具变量估计来描述两个重要的检验。

内生性检验

483　当解释变量是外生的时，2SLS 估计量不如 OLS 有效；正如我们已看到的，2SLS 估计值会有非常大的标准误。因此，检验一个解释变量的内生性是有用的，它说明了 2SLS 甚至是否必要。获取这样的检验相当简单。

举例说明，假定我们有单一的被怀疑的内生变量

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \tag{15.49}$$

式中，z_1 和 z_2 是外生的。我们有另外两个外生变量 z_3 和 z_4，它们不出现在式 (15.49) 中。如果 y_2 与 u_1 不相关，我们该用 OLS 估计式 (15.49)。对此我们如何检验呢？豪斯曼 (Hausman, 1978) 建议直接比较 OLS 和 2SLS 估计值，判断其差异是否在统计上显著。毕竟，如果所有变量外生，OLS 和 2SLS 都是一致性的。如果 2SLS 与 OLS 的差异显著，我们断定 y_2 必定是内生的（z_j 保持外生性）。

计算 OLS 和 2SLS，看估计值是否实际上有差异，这是个好主意。为了判断差异是否在统计上显著，用回归来检验更容易。这是以估计 y_2 的诱导

型为基础的，此时诱导型为

$$y_2=\pi_0+\pi_1 z_1+\pi_2 z_2+\pi_3 z_3+\pi_4 z_4+v_2 \tag{15.50}$$

现在，因为各个 z_j 与 u_1 不相关，所以 y_2 与 u_1 不相关当且仅当 v_2 与 u_1 不相关；这是我们希望检验的。写成 $u_1=\delta_1 v_2+e_1$，其中 e_1 与 v_2 不相关，且有零均值。那么，u_1 与 v_2 不相关当且仅当 $\delta_1=0$。检验这一点最容易的方法是将 v_2 作为添加的回归元包括在式（15.49）中，做 t 检验。这么做惟一的问题是：v_2 不能被观测到，因为它是式（15.50）中的误差项。可是因为我们能用 OLS 估计 y_2 的诱导型，我们可以获取诱导型残差 $\hat{v}_2$。因此，我们用 OLS 估计

$$y_1=\beta_0+\beta_1 y_2+\beta_2 z_1+\beta_3 z_2+\delta_2\hat{v}_2+\text{误差项} \tag{15.51}$$

并用 t 统计量检验 H_0：$\delta_1=0$。如果以一个小的显著水平拒绝 H_0，我们因 v_2 与 u_1 相关推断出 y_2 是内生的。

例 15.7　职业女性的教育回报

通过仅利用职业女性的数据估计诱导型方程（15.41），从中获得残差 $\hat{v}_2$，并将它们包括在方程（15.40）中，我们可以检验方程（15.40）中 educ 的内生性。当我们这么做时，$\hat{v}_2$ 的系数 $\hat{\delta}_1=0.058$，且 $t=1.67$。它是 u_1 与 v_2 之间适度正相关的证据。同时报告两个估计值也许是个好主意，因为教育回报的 2SLS 估计值（6.1%）远低于 OLS 估计值（10.8%）。

单一解释变量的内生性检验

484　（i）通过将 y_2 对所有的外生变量（包括那些在结构方程中和另外的 IV）做回归估计它的诱导型，获得残差 $\hat{v}_2$。

（ii）把 $\hat{v}_2$ 添加到结构方程中（其包括了 y_2），用 OLS 回归检验 $\hat{v}_2$ 的显著性。如果 $\hat{v}_2$ 的系数在统计上异于零，我们推断出 y_2 确实是内生的。我们也许要用对异方差性强健的 t 检验。

第（ii）部分中的回归有一个有趣的特点，所有变量（除了 $\hat{v}_2$）的估计值等同于 2SLS 估计值。例如，用 OLS 估计方程（15.51）所给出的 $\hat{\beta}_j$ 等同于方程（15.49）中的 2SLS 估计值。这是一个简单的检查方法，得以看出你是否在内生性检验中做了正确的回归。它还对 2SLS 给出了另一个解释：在方程（15.51）的 OLS 回归中把 $\hat{v}_2$ 包括进来而清理出了 y_2 的内生性（指把 y_2 的内生部分清理了出来——译者注）。

我们还可以检验复解释变量的内生性。对于每个被怀疑的内生变量，我们如第（i）部分那样获得诱导型残差。然后，我们用 F 检验在结构方程中检验这些残差的联合显著性。联合显著性表明至少有一个被怀疑的解释变量是内生的。被检验的排斥性约束的数目就是被怀疑的内生解释变量的数目。

检验过度识别约束

当我们在15.1节中介绍简单的工具变量估计量时，我们强调IV必须满足两个必需条件：它必须与误差不相关；与内生解释变量相关。我们在相当复杂的模型中已看到，如何判断在诱导型回归中是否能用一个 t 或 F 检验来检验第二个必需条件。我们声称第一个必需条件不能被检验，因为它涉及IV与未观测到的误差之间的相关。然而，如果我们有不止一个的工具变量，我们就能有效地检验它们中的一部分是否与结构误差不相关。

作为一个例子，在有另外两个工具变量 z_3 和 z_4 的条件下，重新考虑方程（15.49）。我们知道仅用 z_3 作为 y_2 的IV，就能估计方程（15.49）。给定IV估计值，我们就能计算残差 $\hat{u}_1=y_1-\hat{\beta}_0-\hat{\beta}_1y_2-\hat{\beta}_2z_1-\hat{\beta}_3z_2$。因为 z_4 在估计中根本没用到，我们可以验证 z_4 与 $\hat{u}_1$ 在样本中是否相关。如果它们相关，z_4 不是 y_2 的有效IV。当然，这并没有告诉我们 z_3 与 u_1 是否相关；实际上，因为它是个有用的检验，我们必须假定 z_3 与 u_1 不相关。然而，如果 z_3 和 z_4 是用相同的逻辑来选择的——例如母亲的教育和父亲的教育——发现 z_4 与 u_1 相关将使人对用 z_3 作为IV产生怀疑。

因为 z_3 和 z_4 的角色可以交换，若是假定 z_4 与 u_1 不相关，我们也可以检验 z_3 与 u_1 是否相关。我们该用哪个检验呢？结果是，我们对检验的选择是无关紧要的。我们必须假定至少有一个IV是外生的。然后，我们可以对2SLS中所用的**过度识别约束**（overidentifying restrictions）进行检验。根据我
485 们的用意，过度识别约束的数目简单地就是额外的工具变量的数目。假定我们只有一个内生解释变量。如果我们只有 y_2 的单一个IV，而没有过度识别约束，也就没什么可检验的。如果我们有 y_2 的两个IV，如同前面的例子中那样，则我们有一个过度识别约束。如果我们有三个IV，则有两个过度识别约束，等等。

检验过度识别约束是相当简单的。我们必须获得2SLS残差，然后做一个辅助回归。

检验（任意多个）过度识别约束

（i）用2SLS估计结构方程，获得2SLS残差 $\hat{u}_1$。

（ii）将 $\hat{u}_1$ 对所有外生变量回归，获得 R^2，即 R_1^2。

（iii）在所有IV都与 u_1 不相关的虚拟假设下，$nR_1^2 \overset{a}{\sim} \chi_q^2$，其中 q 是模型之外的工具变量的数目减去内生解释变量的总数目。如果 nR_1^2 超过了 χ_q^2 分布中的（例如）5%临界值，我们拒绝 H_0，并推断出至少部分的IV不是外生的。

例15.8　职业女性的教育回报

当我们在式（15.40）中用用 *motheduc* 和 *fatheduc* 作为 *educ* 的IV，有

一个过度识别约束。将 2SLS 残差 $\hat{u}$ 对 *exper*，*exper*^{2}，*motheduc* 和 *fatheduc* 做回归，得出 $R_1^2=0.000\,9$。因此，$nR_1^2=428\ (0.000\,9)=0.385\,2$，这在 χ_1^2 分布中是一个非常小的值（p 值 $=0.535$）。因此，父母亲的教育变量通过了过度识别检验。当我们将丈夫的教育增添到 IV 表中，我们得到两个过度识别约束，$nR_1^2=1.11$（p 值 $=0.574$）。因此，将 *huseduc* 增添到 IV 表中似乎是合理的，因为它减少了 2SLS 估计值的标准误：运用所有三个工具，得出的 *educ* 的 2SLS 估计值为 0.080（se $=0.022$），这使得 *educ* 比不用 *huseduc* 作为 IV 时（$\hat{\beta}_{educ}=0.061$，se $=0.031$）显著得多。

在前面的例子中，我们提到了关于 2SLS 的一个普遍事实：在标准的 2SLS 假定下，在表中增添变量提高了 2SLS 的渐近有效性。可是要求任何新的工具实际上是外生的——否则，2SLS 将甚至不是一致性的——而且这只是个渐近的结果。在具备有典型样本容量可供使用的条件下，增添过多的工具——即增加过度识别约束的数目——会导致 2SLS 中的严重偏误。详细的
486 讨论将使我们大大偏离正题。邦德、杰埃格和贝克（Bound，Jaeger and Baker，1995）给出了一个好的例子，他们认为安格里斯特和克鲁格（Angrist and Krueger，1991）用许多工具变量获得的教育回报的 2SLS 估计值，很可能是严重有偏的。（即使有成百上千的观测值！）

无论何时我们有多于所需的工具，都可以用过度识别检验。如果我们有恰好足够的工具，该模型称为是恰好识别的，第（ii）部分中的 R^2 将恒等于零。正如我们前面提到的，在恰好识别情况下我们不能检验工具的外生性。

可以使检验成为对任意形式的异方差性强健的 t 检验。关于细节，参见 Wooldridge（1999，第 5 章）。

15.6 异方差性条件下的 2SLS

2SLS 中的异方差性提出了本质上与 OLS 情况下相同的问题。最重要的是有可能对任意和未知形式的异方差性获得（渐近）强健的标准误和检验统计量。一些软件包按常规做此检验。

我们也能用类似布劳殊-培甘检验的方法来检验异方差性，对此我们在第 8 章中讨论过了。令 $\hat{u}$ 表示 2SLS 残差，并且令 z_1，z_2，…，z_m 表示所有的解释变量（包括那些用做内生解释变量的 IV）。那么，在合理的假定下，［例如，在 Wooldridge（1999，第 5 章）中已有详细的说明］，一个渐进有效的统计量是在 $\hat{u}^2$ 对 z_1，z_2，…，z_m 的回归中用于检验联合显著性的通常的 F 统计量。如果诸 z_j 是联合显著的，同方差性的虚拟假设就被拒绝。

如果我们将它应用于例 15.8 中，用 *motheduc*、*fatheduc* 和 *huseduc* 作为

educ 的工具，我们得到 $F_{5,422}=2.53$，p 值 $=0.029$。这是在5%水平上异方差性的证据。我们也许想要算出这个对异方差性强健的标准误来说明这一点。

如果我们知道误差的方差是如何依赖于外生变量的，就可以用加权的2SLS方法，它本质上与8.4节中是一样的。在估计出 $\mathrm{Var}(u|z_1, z_2, \cdots, z_m)$ 的一个模型之后，我们将第 i 次观测的因变量、解释变量和所有工具变量除以 $\sqrt{\hat{h}_i}$，其中 $\hat{h}_i$ 表示估计的方差。（常数既是解释变量又是IV，将它除以 $\sqrt{\hat{h}_i}$；参见8.4节。）这样，我们用变换了的工具在变换了的方程中应用2SLS。

15.7 2SLS应用于时间序列方程

当我们将2SLS应用于时间序列数据时，在第10章、第11章和第12章中对OLS所作的许多考虑都是适用的。写出每个时期的结构方程为

$$y_t=\beta_0+\beta_1x_{t1}+\cdots+\beta_kx_{tk}+u_t \tag{15.52}$$

其中也许一个或多个解释变量 x_{tj} 与 u_t 相关。用 $z_{t1}, \cdots, z_{tm}$ 表示一组外生变量：

$$\mathrm{E}(u_t)=0, \mathrm{Cov}(z_{tj}, u_t)=0, \quad j=1, \cdots, m$$

任何一个外生解释变量都同时是一个 z_{tj}。为了识别，$m\geqslant k$ 是必须的（我们至少有与解释变量一样多的外生变量）。

487 对时间序列或横截面数据，2SLS的作用是完全相同的，只是对于时间序列数据，2SLS的统计性质依赖于基本序列的趋势性的和相关的性质而已。特别是，如果我们有趋势性的因变量或解释变量，我们必须小心地把趋势包括进来。由于时间趋势是外生的，它总是可以作为自身的工具变量来用。如果运用了月份或季度数据，季节虚拟变量同样可作为工具。

具有强持续性的序列（有单位根）必须谨慎使用，正像OLS的情况那样。往往在估计之前要对方程进行差分，包括对工具变量进行差分。

在类似第11章中OLS渐近性质的假定下，运用时间序列数据的2SLS是一致性的，且渐近地服从正态分布。实际上，如果我们在表述假定时用工具变量代替解释变量，我们只需要对2SLS增添识别假定。例如，同方差性的假定可表述为

$$\mathrm{E}(u_t^2|z_{t1}, \cdots, z_{tm})=\sigma^2 \tag{15.53}$$

无序列相关的假定可表述为

$$\mathrm{E}(u_tu_s|z_t, z_s)=0, \quad t\neq s \tag{15.54}$$

式中，z_t 为时间 t 的所有外生变量。关于假定的详尽表述在本章附录中给

出。我们将在第 16 章中为时间序列问题提供 2SLS 的例子，并参见习题 15.5。

问题 15.4 ☞ 检验政府支出增长对产出增长的影响的一个模型是

$$gGDP_t = \beta_0 + \beta_1 gGOV_t + \beta_2 INVRAT_t + \beta_3 gLAB_t + u_t$$

式中，g 为增长率；*GDP* 为实际国内生产总值；*GOV* 为实际政府支出；*INVRAT* 为国内投资总额对 *GDP* 的比率；*LAB* 为劳动力的多少 ［参见 Ram（1986）中的方程(6)。］表示 $t-1$ 年的总统是否是共和党人的一个虚拟变量，在什么假定下可以是 $gGOV_t$ 的适宜的 IV?

如同在 OLS 的情况下那样，时间序列数据往往违背了无序列相关的假定。幸运的是，检验 AR（1）序列相关是十分简单的。如果我们写出 $u_t=\rho u_{t-1}+e_t$，并将之代入方程（15.52），我们得到

$$y_t=\beta_0+\beta_1 x_{t1}+\cdots+\beta_k x_{tk}+\rho u_{t-1}+e_t,\quad t\geqslant 2 \tag{15.55}$$

为检验 H_0：$\rho=0$，我们必须用 2SLS 残差 $\hat{u}_{t-1}$ 代替 u_{t-1}。此外，如果 x_{tj} 在式（15.52）中是内生的，那么它在式（15.55）中也是内生的，所以我们仍需要运用一个 IV。因为 e_t 与 u_t 的所有过去值不相关，$\hat{u}_{t-1}$ 可用做它自身的工具。

在 2SLS 之后检验 AR（1）序列相关

（i）用 2SLS 估计式（15.52），获得 2SLS 残差 $\hat{u}_t$。

（ii）用 2SLS 估计

488

$$y_t=\beta_0+\beta_1 x_{t1}+\cdots+\beta_k x_{tk}+\rho\hat{u}_{t-1}+error_t,\quad t=2,\cdots,n$$

其中除 $\hat{u}_{t-1}$ 之外运用了与第（i）部分中相同的工具。用 $\hat{\rho}$ 的 t 统计量检验 H_0：$\rho=0$。

如同在第 12 章中得出的 OLS 形式的该检验那样，t 统计量仅具有渐近的正确性，但实际中它往往会很适用。可用对异方差性强健的形式来防止异方差性。此外，还可把滞后残差增添到方程中，以便运用一个联合的 F 检验来对序列相关的高阶形式进行检验。

如果我们侦察出序列相关会怎么样？一些计量经济学软件包将计算标准误，它们对形式相当一般的序列相关和异方差性是强健的。如果由你的计量经济学软件包做此计算，这是一个妥当而简单的方法。其计算与 12.5 节中 OLS 的那些非常相似。参见 Wooldridge（1995）中的公式和其他计算公式。

另一个选择是运用 AR（1）模型，纠正序列相关。程序与 OLS 的相似，对工具变量提供另外的约束。准差分方程与方程（12.32）中的相同：

$$\tilde{y}_t=\beta_0(1-\rho)+\beta_1\tilde{x}_{t1}+\cdots+\beta_k\tilde{x}_{tk}+e_t,\quad t\geqslant 2 \tag{15.56}$$

式中，$\tilde{x}_{tj}=x_{tj}-\rho x_{t-1,j}$。（正如在 12.3 节中，可以使用 $t=1$ 时的观测值，但为了简单我们在此略去了。）问题是：我们能用什么作为工具变量呢？运

用准差分工具 $\tilde{z}_{tj} = z_{tj} - \rho z_{t-1,j}$ 似乎是自然的。然而，只有式（15.52）中的原误差 u_t 在时间 t，$t-1$ 和 $t+1$ 与工具不相关，这才是可行的。那就是说，工具变量在式（15.52）中必须是严格外生的。例如，这排除了滞后因变量作为 IV 的可能性。它也排除了这样的情况：IV 中将来的变动反映了误差 u_t 中现在和过去的变化。

带有 AR（1）误差的 2SLS

（i）用 2SLS 估计式（15.52），获得 2SLS 残差 $\hat{u}_t$（$t=1, 2, \cdots, n$）。

（ii）从 $\hat{u}_t$ 对 $\hat{u}_{t-1}$（$t=2, \cdots, n$）的回归中获得 $\hat{\rho}$，构造准差分变量 $\tilde{y}_t = y_t - \hat{\rho} y_{t-1}$，$\tilde{x}_{tj} = x_{tj} - \hat{\rho} x_{t-1,j}$ 和 $\tilde{z}_{tj} = z_{tj} - \hat{\rho} z_{t-1,j}$（$t \geqslant 2$）。（记住，在大多数情况下一些 IV 也将是解释变量。）

（iii）运用 $\tilde{z}_{tj}$ 作为工具，用 2SLS 估计式（15.56）（其中 ρ 用 $\hat{\rho}$ 代替）。假定式（15.56）满足本章附录中的 2SLS 假定，通常的 2SLS 检验统计量是渐近生效的。

15.8 2SLS 应用于混合横截面和综列数据

将工具变量方法应用于独立混合的横截面数据并没有产生任何新的问题。正如用 OLS 来估计的模型那样，我们应当常常把时期虚拟变量包括进来，以便考虑加总时间效应。这些虚拟变量是外生的——因为时间的流逝是外生的——从而它们充当了自身的工具。

例 15.9 教育对生育率的影响

489 在例 13.1 中，我们利用 FERTIL1.RAW 中的混合横截面数据，控制其他各种因素，估计教育对女性生育率的影响。正如在桑德（Sander, 1992）的研究中，我们考虑 *educ* 在方程中为内生的可能性。用母亲和父亲的教育水平（*meduc*, *feduc*）作为 *educ* 的工具变量。β_{educ} 的 2SLS 估计值是 -0.153（se = -0.039），相比之下，OLS 估计值为 -0.128（se = -0.018）。2SLS 估计值表明的教育对生育率的影响要稍微大一些，而 2SLS 的标准误超过 OLS 标准误的 2 倍。（实际上，基于 2SLS 的 95% 置信区间很可能包含了 OLS 估计值。）β_{educ} 的 OLS 和 2SLS 估计值在统计上无差异，这一点通过 15.5 节中 *educ* 的内生性检验可以看出来：当把诱导型残差 $\hat{v}_2$ 和表 13.1 中的其他回归元（包括 *educ*）一起包括进来，它的 t 统计值是 0.702，这在任何合理的水平下都不是显著的。因此，在这种情况下，我们得出结论，2SLS 与 OLS 之间的差异源自于抽样误差。

在存在未观测到的效应，且一个或多个随时间而变的解释变量具有内生

性的情况下，工具变量估计可与综列数据，特别是一阶差分相结合，用于参数的一致性估计。以下例子说明了这种方法上的结合。

例 15.10　职业培训与工人生产力

假定我们要估计再受一个小时职业培训对工人生产力的影响。对 1988 年和 1989 年两年，考虑简单的综列数据模型

$$\log(scrap_{it}) = \beta_0 + \delta_0 d88_t + \beta_1 hrsemp_{it} + a_i + u_{it}, \quad t = 1,2$$

式中，$scrap_{it}$ 为公司第 t 年的废料率；$hrsemp_{it}$ 为每位雇员所受职业培训的小时数。通常，我们有不同年份的截距，且未观测到的公司影响 a_i 为常数。

由于 13.2 节中讨论的原因，我们也许担心 $hrsemp_{it}$ 与 a_i 相关，后者包含了未观测到的工人的能力。与前面一样，我们通过差分来消去 a_i：

$$\Delta\log(scrap_i) = \delta_0 + \beta_1\Delta hrsemp_i + \Delta u_i \tag{15.57}$$

正常情形下，我们将用 OLS 来估计该方程。但如果 Δu_i 与 $\Delta hrsemp_i$ 相关，会怎样呢？例如一个厂家也许在雇用更多熟练工人的同时，减低在职培训的时数。这时我们需要一个取代 $\Delta hrsemp_i$ 的工具变量。通常，此类 IV 可能会难以找到，但我们可以利用某些公司在 1988 年得到了职业培训许可这个事实。如果我们假定许可的指派与 Δu_i 不相关——这合情合理，因为授予许可 490 是在 1988 年初——那么倘若 $\Delta hrsemp$ 与 $\Delta grant$ 相关，$\Delta grant_i$ 就是一个有效的 IV。利用 JTRAIN.RAW 中 1987 年与 1988 年之间的差分数据，第一阶段回归为

$$\begin{aligned}\widehat{\Delta hrsemp} &= 0.51 + 27.88\Delta grant\\ &\quad\ (1.56)(3.13)\\ &\quad\ n = 45, R^2 = 0.392\end{aligned}$$

这证实了每位雇员所受职业培训的小时数的变化与 1988 年得到职业培训许可有强的正关联。事实上，得到职业培训许可，使每位雇员所受的培训增加约 28 小时。许可的指派约占 $\Delta hrsemp$ 变异中的 40%。做式（15.57）的两阶段最小二乘估计，得到

$$\begin{aligned}\Delta\log(\widehat{scrap}) &= -0.033 - 0.014\Delta hrsemp\\ &\quad\ (0.127)\quad (0.008)\\ &\quad\ n = 45, R^2 = 0.016\end{aligned}$$

这意味着，每位工人所受职业培训增加 10 小时，估计废料率减少约 14%。对于样本中的公司，1988 年平均的职业培训时间约为每个工人 17 小时，最低为零，最高为 88 小时。

相比之下，式(15.57)的 OLS 估计得出 $\hat{\beta}_1 = -0.0076$ (se $= -0.0045$)，β_1 的 2SLS 估计值在大小上差不多是它的 2 倍，且在统计上稍微显著一些。

当 $T \geqslant 3$ 时，差分方程可能包含序列相关。15.7 节中得出的对 AR（1）

序列相关的检验和纠正同样可以用，其中所有的回归都是对 i 和 t 混合的。

未观测到的效应的模型中含有滞后因变量，也需要 IV 方法来获得一致性估计。原因在于，进行了差分之后，由于 $y_{i,t-1}$ 与 $u_{i,t-1}$ 相关，从而 $\Delta y_{i,t-1}$ 与 Δu_{it} 相关。我们可用 y 的两期或更长的滞后作为 $\Delta y_{i,t-1}$ 的 IV。[细节参见 Wooldridge（1999，第 11 章）。]

差分后的工具变量同样可用于配对样本中。阿申费尔特和克鲁格（Ashenfelter and Krueger，1994）在孪生子之间对工资方程进行差分，以消除未观测到的能力：

$$\log(wage_2) - \log(wage_1) = \delta_0 + \beta_1(educ_{2,2} - educ_{1,1}) + (u_2 - u_1)$$

式中，$educ_{1,1}$ 为由孪生长子报告的其所受的学校教育的年数；$educ_{2,2}$ 为由孪生次子报告的其所受的学校教育的年数。为了解释在测量自我报告的学校教育中可能存在的测量误差，阿申费尔特和克鲁格运用（$educ_{2,1} - educ_{1,2}$）作为（$educ_{2,2} - educ_{1,1}$）的 IV，其中 $educ_{2,1}$ 是由孪生长子报告的、孪生次子所受学校教育的年数，$educ_{1,2}$ 是由孪生次子报告的、孪生长子所受学校教育的年数。β_1 的 IV 估计值是 0.167（$t = 3.88$）。相比之下，一阶差分的 OLS 估计值为 0.092（$t = 3.83$）[参见 Ashenfelter 和 Krueger（1994，表 3）]。

▶ 小结

491 在第 15 章中，我们介绍了工具变量的方法。当一个或多个解释变量为内生时，它可对线性模型中的参数进行一致性估计。工具变量必须具备两个性质：（1）它必须是外生的，即与结构方程中的误差项不相关；（2）它必须与内生解释变量偏相关。寻找具有这两个性质的变量常常富有挑战性。

两阶段最小二乘的方法例行用于经验社会科学中，它容许工具变量数比我们的解释变量更多。一旦运用恰当，在存在内生解释变量的情况下，我们可以估计出其他条件不变的效应，这对横截面、时间序列和综列数据上的应用来说都是对的。但是当工具是低劣的——这意味着它们与误差项相关，而仅与内生解释变量弱相关，或与两者都是弱相关——那么 2SLS 可能还不如 OLS。

当具有有效的工具变量，我们可以用 15.5 节中的检验来检验一个解释变量是否内生。此外，尽管无法检验是否所有的 IV 都是外生的，但我们至少可以检验它们中的一部分——假定我们具有的工具数比进行一致性估计所需要的更多（即模型被过度识别）。运用类似于外生解释变量的模型所用的方法，我们可以检验并处理异方差性和序列相关。

在本章中，我们用遗漏变量和测量误差说明了工具变量的方法。IV 法在联立方程模型中也是必不可少的。这一点我们将在第 16 章中讨论。

关键术语

内生解释变量
含误差变量
排斥性约束
外生解释变量
外生变量
识别
工具变量
工具变量（IV）估计量
自然实验
遗漏变量
阶条件
过度识别约束
秩条件
诱导型方程
结构方程
两阶段最小二乘（2SLS）估计量

习　题

15.1　考虑一个简单的模型来估计，一所大规模的公立大学中个人计算机（PC）所有权对大四毕业生的平均积分点的影响：

$$GPA = \beta_0 + \beta_1 PC + u$$

式中，PC 为二值变量，表示计算机所有权。

(i) 为什么计算机所有权可能与 u 相关？

492 (ii) 解释为什么 PC 很可能与父母的年收入相关联。这是否意味着父母的收入是 PC 的一个好的 IV？为什么是或为什么不是？

(iii) 假定四年前该大学资助大约一半的新生购买计算机，而得到的学生是随机选出的。细致地说明你如何运用这一信息去构造 PC 的一个工具变量。

15.2　假定如同例 6.3 那样你想要估计出勤率对学生成绩的影响。一个基本模型是

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u$$

式中变量的定义与第 6 章的相同。

(i) 令 $dist$ 为学生住宿区到教学大楼的距离。你认为 $dist$ 与 u 不相关吗？

(ii) 假定 $dist$ 与 u 不相关，要成为 $atndrte$ 的一个有效的 IV，$dist$ 还必须满足其他什么假定？

(iii) 假定我们如在方程（6.18）中一样增添交互作用项 $priGPA \cdot atndrte$：

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT$$
$$+ \beta_4 priGPA \cdot atndrte + u$$

如果 $atndrte$ 与 u 相关，则一般认为 $priGPA$ 也与 u 相关。什么将是 $priGPA \cdot atndrte$ 的好的IV 呢？［提示：如果 $\mathrm{E}(u \mid priGPA, ACT, dist) = 0$（在 $priGPA$，ACT，$dist$ 均外生时出现），那么 $priGPA$ 和 $dist$ 的任何函数都与 u 不相关。］

15.3 考虑简单回归模型

$$y = \beta_0 + \beta_1 x + u$$

令 z 为 x 的二值工具变量。运用式（15.10），IV 估计量 $\hat{\beta}_1$ 可以写成

$$\hat{\beta}_1 = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

式中，$\bar{y}_0$ 和 $\bar{x}_0$ 是 $z_i = 0$ 的那部分样本中 y_i 和 x_i 的样本平均数；$\bar{y}_1$ 和 $\bar{x}_1$ 是 $z_i = 1$ 的部分样本中 y_i 和 x_i 的样本平均数。该估计量称为分类估计量，它是由沃尔德（Wald，1940）最先提出来的。

15.4 假定你想利用年时间序列数据，对美国某给定州估计州最低限度工资对 18 岁～25 岁之间的年轻人就业（EMP）的影响。一个简单的模型是

$$gEMP_t = \beta_0 + \beta_1 gMIN_t + \beta_2 gPOP_t + \beta_3 gGSP_t + \beta_4 gGDP_t + u_t$$

式中，MIN_t 为以实际美元表示的最低限度工资；POP_t 为 18 岁～25 岁之间的人口；GSP_t 为州生产总值；GDP_t 为美国国内生产总值。前缀 g 表示从 $t-1$ 年到 t 年的增长率，它通常以对数的差分来近似表示。

493 （i）如果我们担心，该州多多少少会以一些（对我们来说）未观测到的却对年轻人就业有影响的因素作为基础来选择最低限度工资，那么 OLS 估计将存在什么问题？

（ii）令 $USMIN_t$ 为美国最低限度工资，它也是以实际价格来衡量的。你认为 $gUSMIN_t$ 与 u_t 不相关吗？

（iii）按照法律，各州的最低限度工资必须至少与全国的最低限度一样。解释这为什么使得 $gUSMIN_t$ 成为 $gMIN_t$ 的一个可能的 IV 备选。

15.5 参考方程（15.19）和方程（15.20）。假定 $\sigma_u = \sigma_x$，因而误差项中的总体变异与 x 中的一样。假定工具变量 z 与 u 轻微相关：$\mathrm{Corr}(z, u) = 0.1$，又假定 z 与 x 具有略微强一些的相关：$\mathrm{Corr}(z, u) = 0.2$。

（i）IV 估计量的渐近偏误是多少？

（ii）x 与 u 之间必须存在多大程度的相关，OLS 才比 2SLS 具有更大的渐近偏误？

15.6（i）在含有一个内生解释变量、一个外生解释变量和一个额外的外生变量的模型中，得出 y_2 的约简型为式（15.26）并代入结构方程（15.22）中，由此得到 y_1 的约简型为

$$y_1 = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + v_1$$

求以 β_j 和 π_j 表示的 α_j。

（ii）求以 u_1，v_2 和参数表示的诱导型残差 v_1。

（iii）你如何进行 α_j 的一致性估计？

15.7 下面一个简单的模型是来测量选校方案对标准化测验成绩的影响的［有兴趣的话，参见 Rouse（1998）］：

$$score = \beta_0 + \beta_1 choice + \beta_2 faminc + u_1$$

式中，*score* 为一次全州范围测验的成绩；*choice* 为一个二值变量，表示一个学生去年是否去所选择的学校上学；*faminc* 为家庭收入。*choice* 的 IV 是 *grant*，即补助给学生用于选校时学费的美元金额。补助金额因家庭收入水平的不同而不同，因此我们在方程中要控制 *faminc*。

（i）尽管方程中有 *faminc*，为什么 *choice* 还可能与 u_1 相关？

（ii）如果在每个收入等级中补助金额是随机分配的，那么 *grant* 是否与 u_1 不相关？

（iii）写出 *choice* 的诱导型方程。我们需要什么而使 *grant* 与 *choice* 偏相关？

（iv）写出 *score* 的诱导型方程。解释它为什么有用？［提示：你如何解释 *grant* 的系数？］

15.8 假定你想检验女子高中的女生数学是否比男女同校的女生要学得
494 好。你有来自美国某州的高中女生的一个随机样本。*score* 是标准化数学测验的成绩。令 *girlhs* 为表示是否就读于女子高中的虚拟变量。

（i）在方程中你应该控制其他什么因素？（你应当能合理地搜集到这些因素的数据。）

（ii）写出 *score* 关于 *girlhs* 和你在第（i）部分中所列的其他因素的方程。

（iii）假定第（ii）部分中父母的支持和动机是误差项内不可测量的因素。它们是否很可能与 *girlhs* 相关？请解释。

（iv）讨论使女生家周围20英里的范围内女子高中的数目成为 *girlhs* 的一个有效的 IV 所需要的假定。

15.9 假定式（15.8）中没有 *skipped* 的好的工具变量候选者。但你有关于学生的两则信息：综合 *SAT* 成绩和以往学期的累积 *GPA*。你会做些什么，如果不是 IV 估计？

15.10 在近来的一篇论文中，埃文斯和斯科韦伯（Evans and Schwab，1995）研究了就读于天主教高中对将来读大学的概率产生的影响。为具体化，令 *college* 为二值变量，如果读大学等于1，否则为0。令 *CathHS* 也为二值变量，如果就读于天主教高中则等于1。一个线性概率模型为

$$college = \beta_0 + \beta_1 CathHS + other\ factors + u$$

其他因素包括性别、种族、家庭收入和父母的教育。

（i）为什么 *CathHS* 可能与 u 相关？

（ii）埃文斯和斯科韦伯拥有关于每个学生在大二时进行的一次标准化测

验的成绩数据。为改进就读于天主教高中的其余条件不变情况下估计值，我们用这些变量能做些什么？

（iii）令 *CathRel* 为二值变量，如果学生是天主教徒则等于 1。讨论它成为前面方程中 *CathHS* 的一个有效的 IV 所需要的两个必要条件。其中哪个可加以检验？

（iv）不足为奇的是，作为天主教徒对是否就读于一所天主教高中有显著的影响。你认为 *CathRel* 作为 *CathHS* 的工具变量是令人信服的吗？

15.11　考虑一个简单的时间序列模型，其中解释变量具有经典的测量误差：

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t^* + u_t \\
x_t &= x_t^* + e_t
\end{aligned}
\qquad (15.58)
$$

式中，u_t 有零均值，且与 x_t^* 和 e_t 不相关。我们只观测 y_t 和 x_t。假定 e_t 有零均值，与 x_t^* 不相关，而且 x_t^* 也有零均值（最后一项假定只是为了简化代数方法）。

495 （i）写出 $x_t^* = x_t - e_t$，将之代入式（15.58）。它表明如果 $\beta_1 > 0$，新方程中的误差项 v_t 与 x_t 负相关。这对于从 y_t 对 x_t 做回归得到 β_1 的 OLS 估计量来说，意味着什么？

（ii）除前面的假定之外，假定 u_t 和 e_t 均与 x_t^* 和 e_t 的所有过去值特别是与 x_{t-1}^* 和 e_{t-1} 不相关。说明 $\mathrm{E}(x_{t-1}v_t) = 0$，其中 v_t 是第（i）部分中模型的误差项。

（iii）x_t 与 x_{t-1} 是否很可能相关？请解释。

（iv）为了得到 β_0 和 β_1 的一致性估计，第（ii）部分和第（iii）部分能为我们提出什么有用的策略？

计算机习题

15.12　本题使用 WAGE2.RAW 中的数据。

（i）在例 15.2 中，用 *sibs* 作为 *educ* 的一个工具，教育回报的 IV 估计值是 0.122。为了使你自己确信，运用 *sibs* 作为 *educ* 的 IV 与仅将 *sibs* 代入以取代 *educ*，并做 OLS 回归的确是不一样的，试将 log(*wage*) 对 *sibs* 进行回归，并解释你的结果。

（ii）变量 *brthord* 是出生的次序（对最先出生的子女，*brthord* 为 1，对第二个出生的子女，其值为 2，等等）。解释为什么 *educ* 可能与 *brthord* 负相关。进行 *educ* 对 *brthord* 的回归，判断是否存在统计上显著的负相关。

（iii）在方程（15.1）中用 *brthord* 作为 *educ* 的一个 IV，报告并解释其结果。

（iv）现在，假定我们将兄弟姊妹的数目作为解释变量包括进工资方程中，这在某种程度上控制了家庭背景：

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 sibs + u$$

假定我们想用 *brthord* 作为 *educ* 的 IV，并假定 *sibs* 是外生的。*educ* 的诱导型为

$$educ = \pi_0 + \pi_1 sibs + \pi_2 brthord + v$$

表述并检验识别的假定。

（v）用 *brthord* 作为 *educ* 的 IV（*sibs* 作为自身的 IV），估计第（iv）部分中的方程。评论 $\hat{\beta}_{educ}$ 和 $\hat{\beta}_{sibs}$ 的标准误。

（vi）运用第（iv）部分中的拟合值 $\widehat{educ}$，计算 $\widehat{educ}$ 与 *sibs* 之间的相关。用该结论解释你在第（v）部分中发现的结果。

15.13 FERTIL2. RAW 中的数据含有 1988 年博茨瓦纳妇女关于孩子

496 数目、学校教育年数、年龄和宗教、经济地位等变量的信息。

（i）用 OLS 估计以下模型：

$$children = \beta_0 + \beta_2 educ + \beta_2 age + \beta_3 age^2 + u$$

并解释估计值。特别是，固定年龄不变，再受一年教育对生育率的影响估计是多少？如果 100 位妇女再受一年教育，她们期望有的孩子数目将减少多少？

（ii）*frsthalf* 是虚拟变量，如果该妇女在上半年内分娩则为 1。假定 *frsthalf* 与第（i）部分中的误差项不相关，说明 *frsthalf* 是 *educ* 的一个合理的 IV 备选。[提示：你需要做一次回归。]

（iii）通过用 *frsthalf* 作为 *educ* 的 IV，估计第（i）部分中的模型。将所估计的教育的影响与第（i）部分中得出的 OLS 估计值进行比较。

（iv）在模型中增添二值变量 *electric*，*tv* 和 *bicycle*。假定它们都是外生的。用 OLS 和 2SLS 来估计方程，并比较 *educ* 的估计系数。解释 *tv* 的系数，以及为什么电视所有权对生育率有负的影响。

15.14 本题使用 CARD. RAW 中的数据。

（i）我们在例 15.4 中所估计的方程可写成

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \cdots + u$$

其他解释变量在表 15.1 中列出。为使 IV 具有一致性，*educ* 的 IV，即 *nearc*4，必须与 u 不相关。*nearc*4 是否会与误差项内的因素相关，例如未观测到的能力？请解释。

（ii）对于数据集中的男性子样本，可以利用 *IQ* 分数。做 *IQ* 对 *nearc*4 的回归以验证平均 *IQ* 分数是否因在四年制大学附近长大而改变。你将得出什么结论？

（iii）将 *IQ* 对 *nearc*4，*smsa*66 及 1966 年地域性虚拟变量 *reg*662，…，*reg*669 进行回归。排除了地理上的虚拟变量之后，*IQ* 是否与 *nearc*4 有关？如何使该答案与你在第（ii）部分中发现的结果相符合。

（iv）从第（ii）和（iii）部分中，对于在 log（*wage*）方程中控制 *smsa*66 和 1966 年地域性虚拟变量的重要性，你将得出什么结论？

15.15 本题使用 INTDER. RAW 中的数据。关于 3 个月的国库券利率

与通货膨胀率（从消费者价格指数中构建）的一个简单方程为

$$i3_t = \beta_0 + \beta_1 inf_t + u_t$$

（i）用OLS估计该方程，为了后面的比较略去时期1。

（ii）一些经济学家认为消费者价格指数错误地测量了真实的通货膨胀 497 率，以致第（i）部分的OLS承受了测量误差偏误。用 inf_{t-1} 作为 inf_t 的IV，重新估计第（i）部分中的方程。β_1 的IV估计值与OLS估计值相比较将如何？

（iii）现在做方程的一阶差分：

$$\Delta i3_t = \beta_0 + \beta_1 \Delta inf_t + \Delta u_t$$

用OLS估计它，并将 β_1 的估计值与前面的估计值相比较。

（iv）在第（iii）部分中的差分方程中，你能否用 Δinf_{t-1} 作为 Δinf_t 的IV？请解释。［提示：Δinf_t 与 Δinf_{t-1} 是否充分地相关？］

15.16　本题使用CARD.RAW中的数据。

（i）表15.1中，教育回报的IV和OLS估计值之间的差异在经济上是重要的。从式（15.32）中得到诱导型残差 $\hat{v}_2$。（参考表15.1，以便将其他变量包括进回归中。）用其检验 $educ$ 是否是外生的；也就是说，判断OLS与IV之间的差异在统计上是否显著。

（ii）增添 $nearc2$ 作为工具，用2SLS估计方程。$educ$ 的系数的变化会很大吗？

（iii）检验第（ii）部分中的单一过度识别约束。

15.17　本题使用MURDER.RAW中的数据。变量 $mrdrte$ 为谋杀率，即每100 000个人中发生谋杀案的数目；变量 $exec$ 为当年和上一年两年中被处决的犯人的总数目；$unem$ 为州失业率。

（i）有多少个州在1991年、1992年和1993年中至少处决了1个犯人？哪个州处决得最多？

（ii）利用1990年和1993年两年的数据，做一个 $mrdrte$ 对 $d93$，$exec$ 和 $unem$ 的混合回归。你对 $exec$ 的系数如何解释？

（iii）仅利用1990—1993年的变化（对总共51个观测值），用OLS估计以下方程

$$\Delta mrdrte = \delta_0 + \beta_1 \Delta exec + \beta_2 \Delta unem + \Delta u$$

并以通常的形式报告结果。现在，处以死刑看起来是否具有威慑作用？

（iv）处决的变化至少可能部分地与预期谋杀率的变化有关，因而 $\Delta exec$ 与第（iii）部分中的 Δu 相关。假定 $\Delta exec_{-1}$ 与 Δu 不相关也许是合情合理的。（毕竟 $\Delta exec_{-1}$ 依赖于三年或更久以前进行的处决数。）将 $\Delta exec$ 对 $\Delta exec_{-1}$ 做回归，看它们是否充分地相关；解释 $\Delta exec_{-1}$ 的系数。

（v）用 $\Delta exec_{-1}$ 作为 $\Delta exec$ 的IV，重新估计第（iii）部分中的方程。假定 $\Delta unem$ 是外生的，你从第（iii）部分中得出的结论将怎样变化？

498　15.18　本题使用PHILLIPS.RAW中的数据。

(i) 例 11.5 中，我们估计了如下形式的预期扩充菲利普斯曲线

$$\Delta inf_t = \beta_0 + \beta_1 unem_t + e_t$$

式中，$\Delta inf_t = inf_t - inf_{t-1}$。用 OLS 估计该方程时，我们假定供给冲击 e_t 与 $unem_t$ 不相关。如果这是错误的，关于 β_1 的 OLS 估计量可作什么解释？

(ii) 假定 e_t 在给定所有过去信息的条件下是不可预期的：E（$e_t | inf_{t-1}$，$unem_{t-1}$，…）=0。解释为什么这使得 $unem_{t-1}$ 成为 $unem_t$ 的一个好的 IV 候选者。

(iii) 将 $unem_t$ 对 $unem_{t-1}$ 做回归。$unem_t$ 与 $unem_{t-1}$ 是否显著相关？

(iv) 用 IV 估计预期扩充菲利普斯曲线。以通常形式报告结果，并将之与例 11.5 中的 OLS 估计值进行比较。

附录 15A

两阶段最小二乘的假定

本附录讨论了一些假定，以使 2SLSA 具有良好的大样本性质。我们首先表述随机抽样情况下对横截面应用的假定。然后，我们讨论应用于时间序列和综列数据还需要增加什么假定。

假定 2SLS.1（对参数为线性）

总体中的模型可写为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

式中，β_0、β_1，…，β_k 为有含义的未知参数（常数）；u 为不可观测的随机误差或随机干扰项。工具变量用 z_j 表示。

假定 2SLS.2（随机抽样）

我们有关于 y，x_j 和 z_j 的一个随机样本。

假定 2SLS.3（外生工具变量）

误差项 u 有零均值，每个 *IV* 都与 u 不相关。

记住任何 u 与 x_j 不相关的都可充当 IV。

假定 2SLS.4（秩条件）

(i) 工具变量之间不存在完全线性关系。(ii) 识别的秩条件成立。

499

在单一内生解释变量的条件下，如同方程（15.42）中的情况，描述秩条件是容易的。令 z_1，…，z_m 表示外生变量，其中 z_k，…，z_m 不出现在结构模型（15.42）中。y_2 的诱导型为

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \cdots + \pi_{k-1} z_{k-1} + \pi_k z_k + \cdots + \pi_m z_m + v_2$$

因此，我们需要 π_k，…，π_m 中至少有一个不为零。这要求至少有一个外生

变量不出现在式（15.42）中（阶条件）。在两个或更多的内生解释变量条件下阐述秩条件需要矩阵代数。[参见 Wooldridge（1999，第 5 章）。]

定理 15A.1

在假定 2SLS.1～2SLS.4 的条件下，2SLS 估计量是一致性的。

假定 2SLS.5（同方差性）

令 z 表示所有工具变量的集合，因而 $\mathrm{E}(u^2|z)=\sigma^2$。

定理 15A.2

在假定 2SLS.1～2SLS.5 的条件下，2SLS 估计量是渐近正态分布的。渐近方差的一致性估计量如方程（15.43）中所给出的，其中 σ^2 用 $\hat{\sigma}^2=(n-k-1)^2\sum_{i=1}^{n}\hat{u}_i^2$ 代替，$\hat{u}_i$ 是 2SLS 残差。

在给出的五个假定下，2SLS 估计量也是最优 IV 估计量。我们在此仅阐述该结果。证明可以在 Wooldridge（1999）中找到。

定理 15A.3

在假定 2SLS.1～2SLS.5 的条件下，2SLS 估计量在用外生变量的线性组合作为工具的一类 IV 估计量中是渐近有效的。

如果同方差性假定不成立，2SLS 估计量仍是渐近正态的，但标准误（及 t 和 F 统计量）需要校正。许多计量经济学软件包例行地做此校正。此外，2SLS 估计量一般地说不再是渐近有效的 IV 估计量。我们这里不去研究更有效的估计量。[参见 Wooldridge（1999，第 8 章）]。

对于时间序列上的应用，我们必须增添一些假定。首先，如同 OLS 条件下，我们必须假定所有序列（包括 IV）弱相关：这保证了大数定律和中心极限定理成立。为了渐近有效性，同时为了使通常的标准误和检验统计量有效，我们必须增添一个无序列相关的假定。 500

假定 2SLS.6（无序列相关）

方程（15.54）成立。

综列数据上的应用需要类似的一个无序列相关的假定。在 15.7 节中已讨论了序列相关的检验和纠正。

第 16 章　联立方程模型

501 我们在前几章说明了如何用工具变量法解决遗漏变量和测量误差这两个内生性问题。从概念上讲，这些问题很容易懂。在遗漏变量的情形中，在估计一个或多个观测到的解释变量在其他条件不变情况下的影响时，我们想保持一个(或不止一个)变量固定不变。在测量误差情形中，我们想估计某些解释变量对 y 的影响，但我们错误地度量了一个或多个变量。在这两种情况下，如果我们能搜集到更好的数据，就能估计利害参数。

解释变量另一种重要的内生性形式是**联立性**(simultaneity)。当一个或多个解释变量与因变量联合被决定时(特别是如以后将看到的那样，通过一个均衡机制一起决定)，就出现了这个问题。在本章中，我们研究估计简单的联立方程模型(SEM)的方法。尽管对 SEM 系统的处理超出了本书的范围，但我们可以探讨那些已被广泛使用的模型。

估计联立方程模型的主要方法是工具变量法。因此，对联立性问题的解决，本质上与 IV 对遗漏变量和测量误差问题的解决别无二致，但设计和解释 SEM 则富有挑战性。因此，我们在 16.1 节先从对联立方程模型的性质和研究范围的讨论开始。在 16.2 节，我们明确在联立方程的一个方程中用 OLS 一般是有偏误和不一致的。

16.3 节对一个两方程的联立模型中的识别和估计进行了一般性的描述，而 16.4 节则简要地探讨了不止两个方程的模型。联立方程模型还用于加总

的时间序列模型的建造，我们在16.5节讨论了在这些模型中出现的一些特殊问题。16.6节则谈及带综列数据的联立方程模型。

16.1 联立方程模型的性质

在利用联立方程模型时最重要的一点是，要记住方程组中的每个方程都具有其他条件不变的因果性解释。因为我们只观察到均衡结果，所以要求我们在构造联立方程模型中的方程时，使用违反现存事实的逻辑。我们必须从
502 潜在和实际结果方面考虑。

SEM的经典例子是某个商品或生产投入(如劳动)的供给和需求方程。为简洁起见，令 h_s 表示农业工人年劳动供给小时数(在县一级水平上度量的)，而令 w 表示向这类工人提供的平均小时工资。一个简单的劳动供给方程就是

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1 \qquad (16.1)$$

式中，z_1 为某个影响劳动供给的可观测变量——如本县制造业的平均工资；误差项 u_1 则包括了影响劳动供给的其他因素。[这些因素中许多都可以观测到，并可以放到方程（16.1）中；为了说明基本概念，我们只包含了这种因素中的 z_1 一个。] 方程（16.1）是**结构方程**（structural equation）的一个例子。此名称源于劳动供给函数是从经济理论推导出来并具有因果性解释的事实。系数 α_1 度量了劳动供给如何随工资的变化而变化：如果 h_s 和 w 都是对数形式，则 α_1 表示劳动供给弹性。典型地讲，我们预期 α_1 为正（尽管经济理论并不排除 $\alpha_1 \leq 0$）。当工资收入的税率改变时，工人愿意工作的小时数也将改变，劳动供给弹性在决定这种变化时很重要。如果 z_1 是制造业工资，那我们预期 $\beta_1 \leq 0$：其他因素不变，如果制造业工资提高，则进入制造业的工人比进入农业的工人多。

在我们画劳动供给曲线时，我们把小时数勾勒成工资的函数，并保持 z_1 和 u_1 不变。z_1 的改变会移动劳动供给线，u_1 的改变也一样。区别在于，z_1 可被观测到，而 u_1 不能。有时，z_1 被称为**可观测的供给移位因子**，而 u_1 被称为**不可观测的供给移位因子**。

方程（16.1）与我们前而研究过的方程有什么不同呢？区别很微妙。当假设（16.1）对所有可能的工资值都成立时，我们一般不能认为工资随各县的横截面数据面外生地变化。如果我们能进行一个实验，其中对各县的一个样本改变农业和制造业的工资水平，并向工人调查以得到劳动供给 h_s，那我们就能用OLS估计方程（16.1）。不幸的是，这不是一个可控制实验。相反，我们必须将农业生产中所用的人时数数据与这两个部门中的平均工资数据一起搜集。在决定如何分析这些数据时，我们必须了解，最好用劳动供给与需求的相互作用来描述。在劳动市场出清的假定下，我们实际上观察到的是工资和工作小时数的**均衡值**。

为了描述均衡工资与均衡小时数的决定，需要引入劳动需求。我们假设劳动需求由下式给出

$$h_d = \alpha_2 w + \beta_2 z_2 + u_2 \tag{16.2}$$

式中，h_d 为需要的小时数。像供给函数一样，我们保持 z_2 和 u_2 不变，在图上将需求的小时数作为工资的一个函数而画出来。变量 z_2（比方说农地面积）是**可观测的需求移动因子**，而 u_2 是**不可观测的需求移动因子**。

恰如劳动供给方程一样，劳动需求方程也是一个结构方程：它可以从农民的利润最大化考虑中得到。如果 h_d 和 w 都是对数形式，则 α_2 就是劳动 503 需求弹性。经济理论告诉我们 $\alpha_2 < 0$。由于劳动和土地在生产中是互补品，所以我们预期 $\beta_2 > 0$。

注意方程（16.1）和方程（16.2）如何描述了完全不同的关系。劳动供给是工人的行为方程，而劳动需求则是农民的行为关系。每个方程都有一个其他条件不变的解释并各自成立。它们在计量经济分析中之所以相互联系，只是因为所观测到的工资和小时数由供给和需求之间的相互作用决定。换句话说，对于每个县 i，所观测到的小时数 h_i 和所观测到的工资 w_i 由下式均衡条件决定：

$$h_{is} = h_{id} \tag{16.3}$$

因为我们只观测到每个县 i 的均衡小时数，所以我们用 h_i 表示所观测到的小时数。

将式（16.3）中的均衡条件与劳动和需求方程合并就得到

$$h_i = \alpha_1 w_i + \beta_1 z_{i1} + u_{i1} \tag{16.4}$$

和

$$h_i = \alpha_2 w_i + \beta_2 z_{i2} + u_{i2} \tag{16.5}$$

其中我们明确地包括下标 i，以强调 h_i 和 w_i 都表示各个县的均衡观测值。这两个方程就构成了一个**联立方程模型**（simultaneous equations model, SEM），它有几个重要特征。首先，给定 z_{i1}，z_{i2}，u_{i1}和 u_{i2}，这两个方程就决定了 h_i 和 w_i。（实际上，我们必须假定 $\alpha_1 \neq \alpha_2$，即供给函数和需求函数的斜率不同；参见习题 16.1。）出于这个原因，h_i 和 w_i 是这个 SEM 中的**内生变量**（endogenous variables）。z_{i1}和 z_{i2}怎么样呢？由于它们在模型外决定，我们把它们看成**外生变量**（exogenous variables）。从统计观点来看，关于 z_{i1} 和 z_{i2}的关键假定是，它们都与供给和需求误差（分别是 u_{i1}和 u_{i2}）无关。由于这些误差出现在结构方程中，所以它们是**结构误差**（structural errors）的例子。

第二个重要的地方是，如果模型中不包括 z_1 和 z_2，就无法分辨哪个是供给方程，哪个是需求方程。当 z_1 表示制造业工资时，经济逻辑告诉我们，它是农业劳动供给中的一个因素，因为它度量了从事农业工作的机会成本；当 z_2 表示农地面积时，生产理论也意味着，它应出现在劳动需求函数中。因此，我们知道式（16.4）代表劳动供给，而式（16.5）代表劳动需求。如果 z_1 和 z_2 相同（如能够同时影响供给和需求的本县成人平均受教育水平），那么这两个方程看起来一样，也就没有指望知道估计的是哪一个。概言之，

这说明了联立方程模型中的识别问题，我们在16.3节将会更一般性地讨论。

504 SEM最令人信服的例子与劳动供给与需求的例子具有相同的特点。每个方程应该有一个行为上的自足的其他条件不变的解释。由于我们只能观测到均衡结果，所以在具体说明一个SEM时，就要求我们问这样一些违反事实的问题：如果工资不同于其均衡值，工人将提供多少劳动？例16.1阐明了这个特征。

例16.1　谋杀率与警察部门的规模

城市通常要决定执法力度要增加到多大才会使该市的谋杀率下降。提出这个方程的一个简单横截面模型是

$$murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + u_1 \qquad (16.6)$$

式中，*murdpc*为人均谋杀数；*polpc*为人均警察数量；*incpc*为人均收入。(因此我们没有包括下标i。)我们取人均收入作为这个方程的外生变量。实践中，我们还将包括诸如年龄和性别分布、受教育水平、地理变量及度量惩罚严厉程度的变量等其他因素。为集中思想，我们只考虑方程(16.6)。

我们希望回答的问题是：如果一个城市外生地增加警力规模，这种增加总体上会降低谋杀率吗？如果我们能对一个城市的随机样本外生地选择警力规模，就可以用OLS估计方程(16.6)。我们当然不能做这样的实验。但我们究竟能否认为警力规模是外生决定的呢？可能不能。一个城市的执法支出至少部分地由其预期谋杀率决定。为了反映这一点，我们阐述第二个关系：

$$polpc = \alpha_2 murdpc + \beta_{20} + \text{其他因素} \qquad (16.7)$$

我们预期$\alpha_2>0$：其他因素不变，预期谋杀率越高的城市，其人均警察数量也越多。一旦我们确定了方程(16.7)中的其他因素，我们就得到了一个两方程的联立方程模型。我们实际上只关心方程(16.6)，但如我们将在16.3节中见到的那样，为了估计第一个方程，我们需要精确地知道第二个方程是如何设定的。

最后，注意到方程(16.6)描述了城市警官的行为，而方程(16.7)则描述了潜在谋杀犯的行为。这就是每个方程明显具有其他条件不变的解释，并使方程(16.6)和方程(16.7)成为一个适当的联立方程模型。

我们接着给出一个不适当使用SEM的例子。

例16.2　住房支出和储蓄

505 假设我们对总体中的一个随机家庭假定其年度住房支出和储蓄由如下两个式子联合决定

$$housing = \alpha_1 saving + \beta_{10} + \beta_{11} inc + \beta_{12} educ + \beta_{13} age + u_1 \qquad (16.8)$$

$$saving = \alpha_2 housing + \beta_{20} + \beta_{21} inc + \beta_{22} educ + \beta_{23} age + u_2 \qquad (16.9)$$

式中，*inc* 为年收入；*educ* 和 *age* 都以年度量。这些方程初看起来像是住房和储蓄支出如何决定的合理方法。但我们必须问一下：其中一个方程在没有另一个方程的情况下有什么价值？由于住房支出和储蓄支出都是由同一个家庭作出选择的，所以没有一个方程具有其他条件不变的解释。问如下问题也就毫无意义：如果 *saving* 外生地变化，对 *housing* 的影响如何？任何一个基于经济学原理的模型（特别是效用最大化），都将使 *housing* 和 *saving* 的最优选择成为 *inc* 及住房和储蓄的相对价格的函数。变量 *educ* 和 *age* 将影响对消费、储蓄和风险的偏好。因此，*housing* 和 *saving* 应该分别成为收入、受教育水平、年龄和其他影响效用最大化问题的变量（如住房和其他方式储蓄不同的回报率）的函数。

即使我们假定方程（16.8）和方程（16.9）中的 SEM 讲得通，也没有办法估计其参数。（我们在 16.3 节更一般地讨论这个问题。）除非我们假定收入、受教育水平或年龄在一个方程出现而在另一个方程不出现（这又讲不通），否则这两个方程无法区别。

尽管这是 SEM 的一个不好的例子，但我们仍可以对如下检验感兴趣：在其他条件不变情况下，住房支出和储蓄之间是否存在取舍。这时我们将只须用 OLS 估计一个方程，比方说方程（16.8），除非有遗漏变量或测量误差的问题。

具有例 16.2 的特点的 SEM 应用太多了。关键的特征是两个方程都表示了同一个经济主体的行为，因此没有一个方程能独自成立。对比之下，例 16.1 中供给和需求的例子则自然具有其他条件不变的解释。在某些情况下由一些简单模型所支持的经济逻辑可以帮助我们明智地使用 SEM（和知道什么时候不用 SEM）。

问题 16.1 平迪克和罗宾费尔德（Pindyck and Rubinfeld, 1992, 11.6 节）描述了一个广告模型，其中垄断企业选择利润最大化的价格和广告支出水平。这意味着我们应该使用一个 SEM 来做这些变量在企业一级上的模型吗？

16.2 OLS 中的联立性偏误

506 有用的是看到，在一个简单模型中，与因变量同时决定的解释变量一般都与误差项相关，这就导致 OLS 中存在偏误和不一致性。我们考虑两个方程的结构模型：

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \tag{16.10}$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2 \tag{16.11}$$

并专门估计第一个方程。变量 z_1 和 z_2 都是外生的，所以每个都与 u_1 和 u_2 无关。为简单起见，我们将每个方程的截距项都去掉了。

为了证明 y_2 通常都与 u_1 相关，我们求解这两个方程，用外生变量和误差项表示 y_2。如果我们将式（16.10）的右边作为 y_1 代入式（16.11）中，得到

$$y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$$

或

$$(1 - \alpha_2\alpha_1)y_2 = \alpha_2\beta_1 z_1 + \beta_2 z_2 + \alpha_2 u_1 + u_2 \tag{16.12}$$

现在，为了解出 y_2，我们必须对参数做一个假定：

$$\alpha_2\alpha_1 \neq 1 \tag{16.13}$$

这个假定是否很有限制性则取决于应用。在例 16.1 中，我们认为 $\alpha_1 \leqslant 0$ 和 $\alpha_2 \geqslant 0$，这就意味着 $\alpha_1\alpha_2 \leqslant 0$；因此式（16.13）对例 16.1 而言很合理。

如果式（16.13）中的条件成立，则我们可以将式（16.12）除以（$1 - \alpha_2\alpha_1$）而将 y_2 写成

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2 \tag{16.14}$$

式中，$\pi_{21} = \alpha_2\beta_1/(1 - \alpha_2\alpha_1)$，$\pi_{22} = \beta_2/(1 - \alpha_2\alpha_1)$ 和 $v_2 = (\alpha_2 u_1 + u_2)/(1 - \alpha_2\alpha_1)$。用外生变量和误差项表示 y_2 的方程（16.14）是 y_2 的**约简型**（reduced form），我们在第 15 章工具变量估计的背景下介绍过这个概念。参数 π_{21} 和 π_{22} 被称为**约简型参数**（reduced parameters）；注意它们是结构方程（16.10）和方程（16.11）中出现的**结构型参数**（structural parameters）的非线性函数。

约简型误差 v_2 是结构型误差 u_1 和 u_2 的线性函数。因为 u_1 和 u_2 都与 z_1 和 z_2 无关，所以 v_2 也与 z_1 和 z_2 无关。因此，我们可用两阶段最小二乘估计所用的 OLS 一致地估计 π_{21} 和 π_{22}（我们下一节将回过来讨论）。此外，约简型参数有时也有直接意义，尽管我们这里只关心对方程（16.10）的估计。

507 在假定（16.13）下，y_1 的约简型也存在；运算与得到方程（16.14）所用的运算相似。它具有与 y_2 的约简型方程相同的特征。

我们可以用方程（16.14）证明，除非在特殊的假定之下，否则对方程（16.10）的 OLS 估计，将导致方程（16.10）中 α_1 和 β_1 的估计量有偏误和不一致。因为根据假定，z_1 和 u_1 不相关，所以问题是 y_2 和 u_1 是否不相关。我们从方程（16.14）中的约简型看出，y_2 和 u_1 相关等价于 v_2 和 u_1 相关（因为假定 z_1 和 z_2 外生）。但 v_2 是 u_1 和 u_2 的一个线性函数，所以它一般都与 u_1 相关。事实上，如果我们假定 u_1 和 u_2 不相关，那么只要 $\alpha_2 \neq 0$，v_2 和 u_1 就一定相关。即便 α_2 等于零（这意味着方程（16.11）中没有出现 y_1），如果 u_1 和 u_2 相关，v_2 和 u_1 也将相关。

当 $\alpha_2 = 0$ 且 u_1 和 u_2 不相关时，y_2 和 u_1 也不相关。这是相当强的要求：如果 $\alpha_2 = 0$，则 y_2 不是与 y_1 同时决定的。如果我们增加 u_1 和 u_2 零相关的假设，那就排除了 u_1 中与 y_2 相关的遗漏变量和测量误差。在这种情况

下，能用 OLS 估计方程（16.10）也就无足为奇了。

当 y_2 与 u_1 因联立而相关时，我们就说 OLS 存在**联立性偏误**（simultaneity bias）。如我们在第 3 章和第 5 章对遗漏变量偏误的讨论中看到的那样，要得到系数中偏误的方向一般都很复杂。但在一些简单的模型中，我们可以决定偏误的方向。比如，假设我们通过去掉 z_1 而简化方程（16.10），并假定 u_1 和 u_2 不相关。于是，y_2 和 u_1 之间的协方差为

$$\begin{aligned}\mathrm{Cov}(y_2, u_1) = \mathrm{Cov}(y_2, v_2) &= [\alpha_2/(1-\alpha_2\alpha_1)]\mathrm{E}(u_1{}^2)\\ &= [\alpha_2/(1-\alpha_2\alpha_1)]\sigma_1{}^2\end{aligned}$$

式中，$\sigma_1^2 = \mathrm{Var}(u_1) > 0$。因此，$\alpha_1$ 的 OLS 估计量中的渐近偏误（或不一致性）与 $\alpha_2/(1-\alpha_2\alpha_1)$ 具有相同的符号。如果 $\alpha_2 > 0$ 和 $\alpha_2\alpha_1 < 1$，则渐近偏误为正。（不幸的是，正如我们在 3.3 节计算遗漏变量偏误一样，这些结论并不能推及更一般的模型，但它们可作为一个有用的指导。）比如在例 16.1 中，我们认为 $\alpha_2 > 0$ 和 $\alpha_2\alpha_1 \leqslant 0$，这意味着 α_1 的 OLS 估计量将有正的偏误。如果 $\alpha_1 = 0$，那么 OLS 将大致估计出更多警察对谋杀率具有正的影响；通常 α_1 的估计量向零衰减。如果我们在方程（16.6）中应用 OLS，则可能低估更大警力的有效性。

16.3 结构方程的识别和估计

如我们在上一节所见，OLS 在用于联立方程组中的一个结构方程时是偏误和不一致的。我们在第 15 章了解到用两阶段最小二乘法解决内生解释变量问题的方法，现在来说明 2SLS 如何用于 SEM。

508 这里 2SLS 的操作与第 15 章中的相似。区别在于，由于对每个内生变量都设定了一个结构方程，所以我们立即能看出是否有足够的 IV 去估计每个方程。我们从对**识别**（identification）问题的讨论开始。

两方程联立模型中的识别

我们在第 15 章提到了识别的概念。当我们用 OLS 估计一个模型时，关键的识别条件是每个解释变量都与误差项无关。如我们在 16.2 节说明的那样，这个重要的条件对 SEM 而言一般不再成立。但如果我们有一些工具变量，就像在遗漏变量或存在测量误差时那样，仍能识别（或一致地估计）一个 SEM 方程中的参数。

我们在考虑一个一般性的两方程 SEM 之前，先从一个简单的供给和需求的例子来得到直觉是有帮助的。把这个系统的均衡形式（即施加条件 $q_s = q_d = q$）写成

$$q = \alpha_1 p + \beta_1 z_1 + u_1 \tag{16.15}$$

或

$$q = \alpha_2 p + u_2 \tag{16.16}$$

为简明起见，令 q 表示在县一级的人均牛奶消费量，令 p 表示在这个县每加仑牛奶的平均价格，而令 z_1 表示牛饲料的价格，并假定它（后一价格）外生于牛奶的供给与需求。这意味着，式（16.15）必须是供给方程，因为牛饲料的价格会改变供给（$\beta_1<0$）但不改变需求。而需求方程则没有包含可观测的需求移动因子。

给定（q，p，z_1）的一个随机样本，能估计这些方程中哪一个呢？即哪个是**可识别方程**（identified equation）呢？结果需求方程（16.16）可识别，而供给方程则不能。通过使用第 15 章我们进行 IV 估计的规则，很容易看出这一点：我们可以用 z_1 作为方程（16.16）中价格的一个 IV。但由于 z_1 出现在方程（16.15）中，所以我们没有供给方程中的价格的 IV。

直觉上讲，需求方程之所以能识别，是因为我们有可观测变量 z_1，它移动了供给方程却又不影响需求方程。对给定的 z_1 的变异并假定没有误差，我们可以像图 16.1 所示那样画出需求曲线。不可观测的需求移动因子 u_2 的出现，使我们估计的需求方程含有误差，但这些估计量都是一致的，只要 z_1 与 u_2 无关。

因为没有外生可观测的因素移动需求曲线，所以就画不出供给方程。有些观测不到的因素移动需求曲线没有什么用；我们需要可观测因素。如果像在劳动需求方程（16.2）那样，我们有一个可观测的外生需求移动因子（比
509 如牛奶需求方程中的收入），那供给方程也将是可识别的。

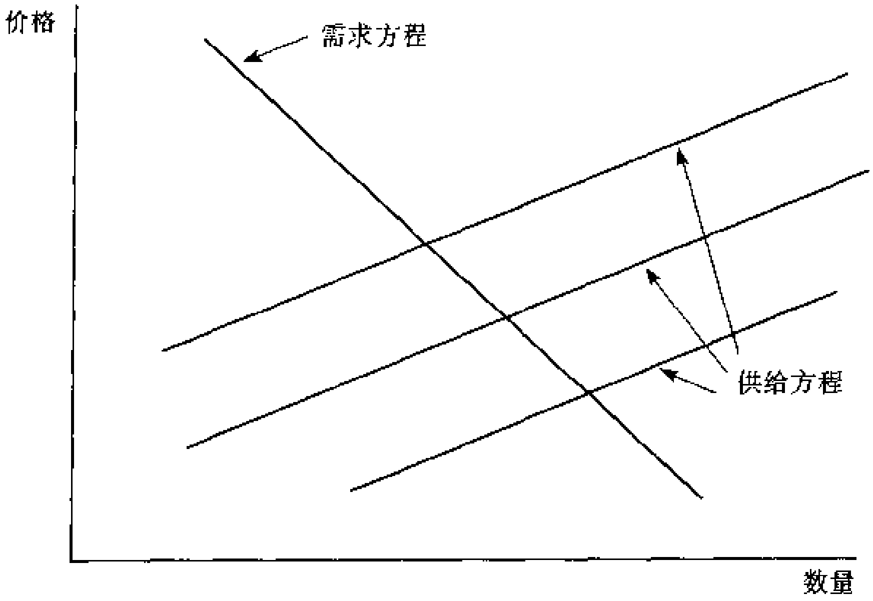


图 16.1 移动供给方程来勾画出需求方程：对外生变量 z_1 的每一个不同值画出一个供给方程

总之，在由方程（16.15）和方程（16.16）构成的系统中，正是供给方

程中一个外生变量的出现，才使我们能估计需求方程。

将识别问题的讨论推广到一般的两方程模型并不困难。把这两个方程写为

$$y_1 = \beta_{10} + \alpha_1 y_2 + z_1\boldsymbol{\beta}_1 + u_1 \tag{16.17}$$

$$y_2 = \beta_{20} + \alpha_2 y_1 + z_2\boldsymbol{\beta}_2 + u_2 \tag{16.18}$$

式中，y_1 和 y_2 为内生变量；u_1 和 u_2 为误差项。第一个方程中的截距是 β_{10}，而第二个方程的截距是 β_{20}。变量 z_1 表示出现在第一个方程中 k_1 个外生变量的集合：$z_1 = (z_{11}, z_{12}, \cdots, z_{1k_1})$。类似地，变量 z_2 表示出现在第二个方程中 k_2 个外生变量的集合：$z_2 = (z_{21}, z_{22}, \cdots, z_{2k_2})$。在许多情况下，$z_1$ 和 z_2 都有重叠部分。作为一个简略形式，我们用符号

510

$$z_1\boldsymbol{\beta}_1 = \beta_{11}z_{11} + \beta_{12}z_{12} + \cdots + \beta_{1k_1}z_{1k_1}$$

和

$$z_2\boldsymbol{\beta}_2 = \beta_{21}z_{21} + \beta_{22}z_{22} + \cdots + \beta_{2k_2}z_{2k_2}$$

即 $z_1\boldsymbol{\beta}_1$ 表示第一个方程中的所有外生变量分别乘上一个系数，$z_2\boldsymbol{\beta}_2$ 也类似。(有些作者用记号 $z_1'\boldsymbol{\beta}_1$ 和 $z_2'\boldsymbol{\beta}_2$。如果你对计量经济学的矩阵代数方法感兴趣，可参见附录 E。)

z_1 和 z_2 一般都包含不同外生变量的事实意味着我们对模型施加了**排除性约束** (exclusion restrictions)。换句话说，我们假定某些外生变量不会出现在第一个方程中，而另有一些则不会出现在第二个方程中。如在前面供给和需求的例子中看到的那样，这就使我们能区分这两个结构方程。

我们什么时候能解方程 (16.17) 和方程 (16.18) 得到 y_1 和 y_2 (作为所有外生变量与结构误差 u_1 和 u_2 的线性函数) 呢？条件与方程 (16.13) 中的条件一样，即 $\alpha_2\alpha_1 \neq 1$。证明与 16.2 节中的简单模型完全相同。在这个假定下，y_1 和 y_2 的约简型存在。

关键问题是，在什么条件下我们能估计比方说方程 (16.17) 中的参数？这是一个识别的问题。识别方程 (16.17) 的秩条件很容易表述。

识别一个结构方程的秩条件

识别（两方程）联立方程模型中第一个方程的充要条件是，第二个方程中至少包含第一个方程所排除的外生变量中的一个（具有非零系数）。

这是识别方程 (16.17) 的充分必要条件。我们在第 15 章讨论的阶条件是秩条件的必要条件。识别第一个方程的阶条件说，这个方程至少要排除一个外生变量。一旦两个方程都设定之后，检查阶条件就很简单。秩条件的要求则多一些：第二个方程中至少有一个第一个方程排除的外生变量，并具有非零的总体系数。这就保证至少有一个外生变量被第一个方程略去，且确实出现在 y_2 的约简型中，所以我们可以用这些变量作为 y_2 的工具变量。如在第 15 章一样，我们可用一个 t 或 F 检验来检验这一点；下面有几个例子。

第二个方程的识别自然与第一个方程的识别如出一辙。同时，如果我们

将方程写成16.1节中劳动供给与需求的例子那样（即 y_1 出现在两个方程中的左边，y_2 出现在两个方程的右边），识别条件是一样的。

例 16.3　已婚工作妇女的劳动供给

为了说明识别问题，考虑已婚劳动妇女的劳动供给。取代需求函数，我们将工资报价写成工作小时数和其他生产力变量的一个函数。加上均衡条件，这两个结构方程为

511

$$hours = \alpha_1 \log(wage) + \beta_{10} + \beta_{11}educ + \beta_{12}age + \beta_{13}kidslt6 + \beta_{14}nwifeinc + u_1 \tag{16.19}$$

和

$$\log(wage) = \alpha_2 hours + \beta_{20} + \beta_{21}educ + \beta_{22}exper + \beta_{23}exper^2 + u_2 \tag{16.20}$$

式中，age 为该妇女以年度量的年龄；$kidslt6$ 为不到6岁的子女的数量；$nwifeinc$ 为该妇女的非工资收入（包括其丈夫的收入）；$educ$ 和 $exper$ 则分别为其受教育的年数和先前工作的年数。假定除 $hours$ 和 $\log(wage)$ 之外所有的变量都是外生的。（这是一个脆弱的假定，因为 $educ$ 可能与任何一个方程中遗漏的能力变量相关。但为说明问题起见，我们忽略遗漏能力变量的问题。）这个系统的函数形式（其中 $hours$ 以水平值形式出现，但 $wage$ 以对数形式出现）在劳动经济学中相当流行。我们可以通过定义 $y_1 = hours$ 和 $y_2 = \log(wage)$ 而将这个系统写成方程（16.17）和方程（16.18）那样。

第一个方程是供给函数。因为劳动供给方程中没有两个外生变量 $exper$ 和 $exper^2$，所以它满足阶条件。这些排除性约束是关键的假定：我们在假定，一旦控制了工资、受教育水平、年龄、幼年子女数和其他收入，过去的工作经验对当前的劳动供给没有影响。人们当然可以对这个假定提出质疑，但我们为说明问题方便而使用它。

给定方程（16.19）和方程（16.20），识别第一个方程的秩条件是，方程（16.20）中的 $exper$ 和 $exper^2$ 至少有一个具有非零系数。若 $\beta_{22} = 0$ 和 $\beta_{23} = 0$，则第二个方程中没有任何第一个方程中也没有出现的外生变量（$educ$ 在两个方程中都出现了）。我们也可用 $\log(wage)$ 的约简型

$$\log(wage) = \pi_{20} + \pi_{21}educ + \pi_{22}age + \pi_{23}kidslt6 + \pi_{24}nwifeinc + \pi_{25}exper + \pi_{26}exper^2 + v_2 \tag{16.21}$$

等价地表述识别方程（16.19）的秩条件。为了识别，我们需要 $\pi_{25} \neq 0$ 或 $\pi_{26} \neq 0$，如我们在第15章讨论过的那样，可以利用一个标准的 F 统计量来检验它。

识别工资报价方程（16.20）的条件是，式（16.19）中的 age、$kidslt6$ 或 $nwifeinc$ 中至少有一个具有非零系数。这就等于假定 $hours$ 的约简型［具有与式（16.21）右边相同的形式］至少取决于 age，$kidslt6$ 或 $nwifeinc$ 中的一个。在设定工资报价方程时，我们假定，一旦解释了工作小时数、受教

育水平和工作经历后，*age*，*kidslt6* 和 *nwifeinc* 对所提供的工资没有影响。如果这些变量对生产力多少有些直接影响，或者妇女因其年龄或年幼子女数量而被歧视时，这就是很不好的假定。

512 在例 16.3 中，我们取所关心的总体为已婚的工作妇女（所以均衡的工作小时数为正），这就排除了已婚但又选择不在家外工作的妇女群。在模型中包括这种妇女会带来一些困难的问题。例如，如果一个妇女不工作，我们就不能观察到她的工资报价。我们在第 17 章会讨论这些问题，但现在，我们必须仅针对 $hours>0$ 的已婚妇女考虑方程（16.19）和方程（16.20）。

例 16.4　通货膨胀与开放度

罗默（Romer，1993）提出了通货膨胀的理论模型，这些模型意味着越开放的国家应该具有越低的通货膨胀率。他的经验分析，用自 1973 年以来进口占国内（或国民）生产总值的比例（这是他对开放度的度量），解释（自 1973 年以来）年平均通货膨胀率。除用 OLS 估计关键方程外，他还使用了工具变量。尽管罗默没有在一个联立方程组中设定两个方程，但他的思想中是一个两方程的方程组：

$$inf = \beta_{10} + \alpha_1 open + \beta_{11}\log(pcinc) + u_1 \tag{16.22}$$

$$open = \beta_{20} + \alpha_2 inf + \beta_{21}\log(pcinc) + \beta_{22}\log(land) + u_2 \tag{16.23}$$

式中，*pcinc* 为美国 1980 年以美元计的人均收入（假定为外生变量）；*land* 为美国以平方英里计的土地面积（也假定为外生变量）。方程（16.22）是我们所关心的方程，并假设 $\alpha_1<0$（越开放的经济通货膨胀率越低）。第二个方程所反映的事实是，开放程度取决于平均通货膨胀率和其他因素。变量 $\log(pcinc)$ 在两个方程中都出现，而假定 $\log(land)$ 仅在第二个方程中出现。其思想是，在其他条件不变的情况下，越小的国家可以越开放（所以 $\beta_{22}<0$）。

利用前面说的识别规则，若 $\beta_{22}\neq 0$，则方程（16.22）可识别。方程（16.23）则因两个外生变量都包含在内而不能识别。但我们感兴趣的是方程（16.22）。

问题 16.2 如果我们有每个国家自 1973 年以来的货币供给增长数据，并假定其为外生变量，这有助于我们识别方程（16.23）吗？

使用2SLS的估计

我们一旦决定了哪个方程被识别，就可以用两阶段最小二乘法去估计它。工具变量由其他方程中出现的外生变量构成。

例 16.5　已婚工作妇女的劳动供给

513 利用 MROZ.RAW 中已婚工作妇女的数据，我们使用 2SLS 估计劳动供给方程（16.19）。全部工具变量集包括 *educ*，*age*，*kidslt*6，*nwifeinc*，*exper* 和 $exper^2$。估计的劳动供给曲线为

$$
\begin{aligned}
\widehat{hours} = {} & \underset{(574.56)}{2\,225.66} + \underset{(470.58)}{1\,639.56}\log(wage) - \underset{(59.10)}{183.75}educ - \underset{(9.38)}{7.81}age \\
& - \underset{(182.93)}{198.15}kidslt6 - \underset{(6.61)}{10.17}nwifeinc \\
& n = 428 \qquad (16.24)
\end{aligned}
$$

它表明劳动供给曲线右上倾斜。log（*wage*）的估计系数具有如下解释：保持其他因素不变，$\Delta \widehat{hours} = 16.4\ (\%\Delta wage)$。通过将这最后一个方程的两边同时乘以 100/*hours*，我们可以计算劳动供给弹性：

$$100(\Delta \widehat{hours}/hours) \approx (1\,640/hours)(\%\Delta wage)$$

或

$$\%\Delta \widehat{hours} \approx (1\,640/hours)(\%\Delta wage)$$

这意味着劳动供给（对工资的）弹性无非就是 1640/*hours*。[在这个模型中，由于 *hours*（不是 log（*hours*））是式（16.24）中的因变量，所以这个弹性不是常数。] 在平均工作小时数 1 303 的水平上，估计的弹性是 1 640/1 303≈1.26，这意味着给定工资 1%的提高，工作小时数的增加大于 1%。这是一个很大的估计弹性。在越高的工作小时数水平上，弹性会越小；而在较低的小时数上，如在 *hours* = 800 时，弹性超过 2。

作为比较，当用 OLS 估计方程（16.19）时，log(*wage*)的系数是 −2.05（se = 54.88），这意味着对工作小时数没有劳动供给效应。为肯定 log(*wage*)实际上在方程(16.19)中是内生的，我们可进行 15.5 节中的检验。当我们在方程中增加约简型残差 $\hat{v}_2$ 并用 OLS 估计时，$\hat{v}_2$ 的 t 统计量就是 −6.61，这是相当显著的，所以 log(*wage*)看起来是内生的。

工资报价方程(16.20)也可以用 2SLS 估计。结果是

$$
\begin{aligned}
\widehat{\log(wage)} = {} & -\underset{(0.338)}{0.656} + \underset{(0.000\,25)}{0.000\,13}hours + \underset{(0.016)}{0.110}\,educ + \underset{(0.019)}{0.035}\,exper \\
& - \underset{(0.000\,45)}{0.000\,71}exper^2 \\
& n = 428 \qquad (16.25)
\end{aligned}
$$

这与前面的工资方程的不同之处在于，将 *hours* 作为一个解释变量包括进来并用 2SLS 解释 *hours* 的内生性（并假定 *educ* 和 *exper* 是外生的）。*hours* 的系数在统计上不显著，这意味着，没有工资报价随着工作小时数的增加而提高的证据。其他系数与我们去掉 *hours* 并用 OLS 估计这个方程时所得到的结果类似。

通过工具变量估计开放度对通货膨胀的影响也很简单。

例 16.6　通货膨胀与开放度

514 在我们利用 OPENNESS.RAW 中的数据估计方程（16.22）之前，我们先检查一下，看 *open* 与所提出的工具变量 log（*land*）之间是否有充分的偏相关。约简型的回归为

$$\begin{aligned}\widehat{open} &= 117.08 + 0.546\log(pcinc) - 7.57\log(land)\\ &\quad\ (15.85)\quad (1.493)\qquad\qquad\quad (0.81)\\ &\quad\ n = 114,\ R^2 = 0.449\end{aligned}$$

log(*land*)的 t 统计量在绝对值上超过了 9，这就证实了罗默关于国家越小越开放的论断。这个回归中 log(*pcinc*)相当不显著的事实无关紧要。

利用 log(*land*)作为 *open* 的一个 IV 估计方程(16.22)，得到

$$\begin{aligned}\widehat{inf} &= 26.90 - 0.337open + 0.376\log(pcinc)\\ &\quad\ (15.40)(0.144)\qquad (2.015)\\ &\quad\ n = 114\end{aligned}\tag{16.26}$$

open 的系数相对单侧对立假设（$\alpha_1 < 0$）约在 1% 的显著性水平上仍是统计显著的。这个影响在经济上也是重要的：进口占 GDP 的份额每提高 1 个百分点，年通货膨胀率约降低 1/3 个百分点。作为对比，其 OLS 估计值为 -0.215（se = 0.095）。

问题 16.3　你如何检验 *open* 的 OLS 和 IV 估计值在统计上是否有差别？

16.4　多于两个方程的联立方程组

联立方程模型也可以由两个以上的方程组成。研究这些模型的一般识别很困难并要用到矩阵代数。一旦一个一般系统中的一个方程被证明是可识别的，那就可以用 2SLS 估计它。

三个或更多方程的系统中的识别问题

我们将用一个三方程的系统来说明在识别复杂的 SEM 中出现的问题。省略掉截距项，将模型写成

$$y_1 = \alpha_{12} y_2 + \alpha_{13} y_3 + \beta_{11} z_1 + u_1 \tag{16.27}$$

$$y_2 = \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2 \tag{16.28}$$

$$y_3 = \alpha_{32} y_2 + \beta_{31} z_1 + \beta_{32} z_2 + \beta_{33} z_3 + \beta_{34} z_4 + u_3 \tag{16.29}$$

式中，y_g 为内生变量；z_j 为外生变量。参数的第一个下标表示方程的编号，
515 第二个则表示变量编号；我们用 α 表示内生变量的参数，β 表示外生变量的
参数。

这些方程中的哪一个可被估计呢？证明含有两个以上方程的 SEM 中的一个方程可识别性一般都很困难，但很容易看出某些方程不能被识别。在方程（16.27）到方程（16.29）中，我们很容易看出，方程（16.29）就不能识别。因为这个方程包含了每一个外生变量，所以找不到 y_2 的工具变量。因此，我们不能一致地估计这个方程的参数。出于我们在 16.2 节讨论的原因，OLS 估计通常都不是一致的。

方程（16.27）怎么样呢？由于 z_2，z_3 和 z_4 都从这个方程中排除，所以看来有可识别的希望——这是**排除性约束**的另一个例子。尽管这个方程有两个内生变量，但我们有 y_2 和 y_3 的三个潜在工具变量。因此，方程通过了阶条件。为完整起见，我们对一般的 SEM 表述阶条件。

识别的阶条件

对任何一个 SEM 中的一个方程，如果它排除的外生变量数不少于其右端包含的内生变量数，那么它就满足识别的阶条件。

第二个方程（16.28）排除了一个外生变量 z_4，而在其右端只有一个内生变量 y_1，所以也通过了阶条件。

如我们在第 15 章和上一节所讨论的那样，阶条件只是识别的必要条件而非充分条件。例如，若 $\beta_{34}=0$，则 z_4 在这个系统中就没有在任何地方出现，这意味着它与 y_1，y_2 或 y_3 都不相关。如果 $\beta_{34}=0$，那么因为 z_4 不能用做 y_1 的一个工具变量，所以第二个方程就不能识别。这再次说明，一个方程的识别取决于其他方程中的参数值（但我们又不可能确切地知道这些参数值）。

还有许多微妙的地方，使得在复杂的 SEM 中识别的阶条件不能成立。为了得到充分条件，我们需要扩展两方程的系统中识别的秩条件，这虽然是可能的，但要用到矩阵代数［比如参见 Wooldridge（1999，第 9 章）］。在许多应用中，人们假定，除非明显不能识别，否则一个满足阶条件的方程就是可以识别的了。

第 15 章有关过度识别和恰好识别方程的术语与 SEM 同时产生。用阶条件的术语来说，方程（16.27）是**过度识别方程**（overidentified equation），因为我们只需要（y_2 和 y_3）两个工具变量，但我们有三个（z_2，z_3 和 z_4）；在这个方程中有一个过度识别约束。一般而言，过度约束个数等于系统中外生变量的总数减去这个方程中解释变量的总数。这一点可以用 15.5 节的过度识别检验来检验。方程（16.28）是一个**恰好识别方程**（just identified equation），而第三个方程是一个**不能识别方程**（unidentified equation）。

估 计

无论一个 SEM 中有多少个方程，每个可识别的方程都可以用 2SLS 估计。某特定方程的工具可由在这个系统中任何地方出现的外生变量组成。对内生性、异方差性、序列相关和过度识别约束的检验可像在第 15 章中那样得到。516

结果证明，对于任何一个由两个或两个以上方程构成的系统，只要被正确设定并符合某些附加假定，系统估计方法一般都比用 2SLS 逐个地估计每一个方程更有效。在 SEM 背景下，最常见的系统估计法是三阶段最小二乘法。无论有没有内生解释变量，这些方法都超出了本书的范围［比如参见 Wooldridge（1999，第 7 章和第 8 章）］。

16.5 利用时间序列的联立方程模型

对 SEM 最早的应用，是估计一个很大的联立方程系统，其中的方程是用来描述一个国家经济系统的。总需求的一个简单凯恩斯模型为（不考虑出口和进口）

$$C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + u_{t1} \tag{16.30}$$

$$I_t = \gamma_0 + \gamma_1 r_t + u_{t2} \tag{16.31}$$

$$Y_t \equiv C_t + I_t + G_t \tag{16.32}$$

式中，C_t 为消费；Y_t 为收入；T_t 为税收收入；r_t 为利率；I_t 为投资；G_t 为政府支出。［比如参见 Mankiw（1994，第 9 章）。］为简洁起见，假定 t 代表年份。

第一个方程是总消费函数，其中消费取决于可支配收入、利率和观测不到的结构误差 u_{t1}。第二个方程是一个很简单的投资函数。方程（16.32）是国民收入会计中的一个恒等式：根据定义就成立，不含误差。因此，估计方程（16.32）没有意义；但为了使模型更完整，我们仍需要这个方程。

由于这个系统中有三个方程，所以一定有三个内生变量。给定前两个方

程，显然我们想让 C_t 和 I_t 成为内生变量。此外，因为这个会计恒等式，所以 Y_t 也是内生的。我们至少在这个模型中将假定 T_t，r_t 和 G_t 是外生的，所以它们与 u_{t1} 和 u_{t2} 不相关。（我们以后将讨论做这种假定的问题。）

如果 r_t 是外生的，那么用 OLS 估计式（16.31）就很自然。但消费函数取决于内生的可支配收入 Y_t。在维持内生假定下，我们有两个工具可用：T_t 和 G_t。因此，根据我们估计横截面方程的惯例，我们就能利用工具（T_t，G_t，r_t）而通过 2SLS 估计式（16.30）。

由于以下几个原因，现在很少估计像式（16.30）到式（16.32）这样的模型。首先，很难在总量的水平上辨明税收、利率和政府支出都外生的假定。税收显然直接取决于收入；比如，第 t 年有单一个边际收入税率 τ_t，则 $T_t=\tau_t Y_t$。我们很容易通过在式（16.30）中以（$1-\tau_t$）Y_t 取代（Y_t-T_t）来表示这种情况，而且如果假定政府支出是外生的，那么我们就能用 2SLS 估计这个方程。如果税率是外生的，也可以把它放到工具变量的名单中。但政府支出和税率真的是外生的吗？如果政府独立于经济系统中发生的情况而确定支出和税率，那它们原则上就可能是外生的。但在现实中，这一点很难
517 实现：政府支出一般都取决于收入水平，而且在高收入水平上，较低的边际税率得到相同的税收收入。此外，假定利率外生极成问题。我们可以设定一个包括货币供给与需求的更现实模型，然后利率就可由 C_t，I_t 和 Y_t 联合决定。但随后找到足够的外生变量来识别这些方程就相当困难（而且这些模型的如下问题仍存在）。

有人认为，政府支出中像国防支出之类的部分［如参见 Hall（1988）和 Ramey（1991）］在大量的联立方程应用中都是外生的。但这并非一致认同，而且无论如何，国防支出也并不总是与内生解释变量适当相关［对此讨论可参见 Shea（1993），而一个例子可参见习题 16.14］。

从式（16.30）到式（16.32）这种模型的第二个问题是，它完全是静态的。特别是对月份或季度数据，但即便是年度数据，我们通常也预期有调整时滞。（支持静态凯恩斯类型模型的一种论点是，它们倾向于描述长期而不担心短期的动态。）容许出现动态并不困难。比如，我们可以在方程（16.31）中增加滞后收入：

$$I_t=\gamma_0+\gamma_1 r_t+\gamma_2 Y_{t-1}+u_{t2} \tag{16.33}$$

换句话说，我们在投资方程中增加了一个**滞后内生变量**（lagged endogenous variable）（但不是 I_{t-1}）。我们可把 Y_{t-1} 作为这个方程中的内生变量吗？在对 u_{t2} 做特殊假定的情况下，回答是肯定的。但我们将一个 SEM 中的滞后内生变量专称为**前定变量**（predetermined variable）。外生变量的滞后值也是前定变量。如果我们假定 u_{t2} 与当前的外生变量（标准的）及所有过去的内生和外生变量都无关，那么 Y_{t-1} 就与 u_{t2} 无关。给定 r_t 的外生性，我们就能用 OLS 估计式（16.33）。

如果在式（16.30）中增加滞后消费，那么在对 u_{t1} 做我们在上一段对 u_{t2} 所做的同样假定下，就可以在这个方程中将 C_{t-1} 看成外生变量。当期可

支配收入仍内生于

$$C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + \beta_3 C_{t-1} + u_{t1} \qquad (16.34)$$

所以我们可利用工具（T_t，G_t，r_t，C_{t-1}）通过 2SLS 估计这个方程；如果投资由式（16.33）决定，则 Y_{t-1} 应加到工具的行列中。［为看出其原因，利用式（16.32）、式（16.33）和式（16.34）求出 Y_t 以外生变量和前定变量 T_t，r_t，G_t，C_{t-1} 和 Y_{t-1} 表示的约简型。因为 Y_{t-1} 出现在这个约简型中，所以它应该可用作一个工具变量。］

在加总的 SEM 中出现动态，至少从预测的角度讲，是对静态 SEM 的一
518 个明显改进。但利用加总时间序列数据估计 SEM 时仍存在一些重要问题（有些我们在第 11 章和第 15 章讨论过）。记得通常的 OLS 和 2SLS 在时间序列应用中的有效性取决于**弱依赖**（weak dependence）的概念。不幸的是，像总消费、收入、投资甚至利率的序列，看起来违背了弱依赖的要求。（用第 11 章的术语，它们有**单位根**。）这些序列还倾向于具有指数趋势，尽管可通过使用对数变换和假定不同的函数形式而部分地克服。通常，不要说小样本，即便是大样本，OLS 和 2SLS 在应用于含有 I（1）变量的方程时的性质都很复杂，并取决于各种假定。我们在第 18 章将简要讨论这些问题。汉密尔顿（Hamilton，1994）给出了更高深而又一般的探讨。

前面的讨论意味着 SEM 不能有效地应用于时间序列数据吗？完全不是。含有趋势和高度持续性的问题，可通过设定一阶差分或增长率的系统而回避。但应该承认，这个系统与以水平值设定的系统不同。［比如，如果我们设定消费的增长是可支配收入增长和利率变化的函数，这就与式（16.30）不同。］我们在前面还讨论过，引进动态并不是特别困难。最后，寻找放进 SEM 中的真正外生变量的问题，对分解数据而言通常容易些。比如，谢伊（Shea，1993）对制造业描述了如何将其他产业的产出（更准确地将是产出增长）用做估计供给方程的一个工具。里梅（Ramey，1991）也包含了通过工具变量估计产业成本函数的一个令人信服的分析，其中利用了时间序列数据。

下面这个例子说明，如何用总量数据检验一个重要的经济理论——消费的持久收入理论，即通常所谓的**持久收入假说**（PIH）。严格地讲，本例中所用到的方法，并非基于联立方程模型，但我们可以认为消费和收入的增长（以及利率）是一起决定的。

例 16.7 对持久收入假说的检验

坎贝尔和曼基（Campbell and Mankiw，1990）利用工具变量方法检验了各种形式的持久收入假说。我们将用 CONSUMP.RAW 中 1959—1995 年的年度数据来模仿其中的一个分析。坎贝尔和曼基使用的是截止到 1985 年的季度数据。

坎贝尔和曼基估计的一个方程（使用我们的符号）是

$$gc_t = \beta_0 + \beta_1 gy_t + \beta_2 r3_t + u_t \tag{16.35}$$

式中，$gc_t \equiv \Delta\log(c_t)$ 为真实人均消费（耐用品除外）的年增长率；gy_t 为真实可支配收入的增长率；$r3_t$ 为以（过去）3 个月期国库券的利率回报度量的真实利率：$r3_t = i3_t - inf_t$，其中通货膨胀率是基于消费者价格指数计
519 算的。消费和可支配收入的增长率是没有趋势和弱依赖的；我们还要假定 $r3_t$ 也是这样，以便我们能应用标准的渐近理论。

方程（16.35）的关键特征是，PIH 意味着，根据在 $t-1$ 时刻或此前所观测到的全部信息，误差项的均值为 0：$\mathrm{E}(u_t|I_{t-1})=0$。但 u_t 不一定与 gy_t 或 $r3_t$ 无关；通常对这个问题的认识是，这些变量是联合决定的，但我们没有写下三个联立方程组。

由于与 $t-1$ 时刻或此前的所有变量都无关，所以估计方程（16.35）的有效工具是 gc，gy 和 $r3$ 的滞后值（以及其他可观测而我们在此又没有使用的变量的滞后值）。我们所关心的假设是什么呢？PIH 的纯粹形式是 $\beta_1=\beta_2=0$。坎贝尔和曼基认为，如果总体中有一定比例的人消费当期收入而不是持久收入，那么 β_1 就为正。真实利率不是常数时 PIH 意味着 $\beta_2>0$。

当我们使用工具 gc_{-1}，gy_{-1}和 $r3_{-1}$通过 2SLS 估计时，我们得到

$$\begin{aligned}\hat{gc}_t &= 0.008\,1 + 0.586 gy_t - 0.000\,27 r3_t \\ &\quad (0.003\,2)(0.135) \quad (0.000\,76) \\ &\quad n=35, R^2=0.678\end{aligned} \tag{16.36}$$

由于 gy 的系数在经济上很大（可支配收入提高 1% 导致消费的增加超过 0.5%），在统计上很显著（$t=4.34$），所以 PIH 的纯粹形式被强烈拒绝。相比之下，真实利率系数则很小并且在统计上不显著。这些发现与坎贝尔和曼基所得到的结论在性质上是一样的。

PIH 还意味着误差 $\{u_t\}$ 是序列相关的。在 2SLS 估计后，我们得到残差 $\hat{u}_t$，并将 $\hat{u}_{t-1}$作为一个附加的解释变量包括在方程（16.36）中，以充当自己的工具变量（参见 15.7 节）。由于 $\hat{u}_{t-1}$的系数为 $\hat{\rho}=0.187$（$\mathrm{se}=0.133$），这表示还是存在着某种正序列相关的证据，尽管在 5% 的显著性水平上还不存在。坎贝尔和曼基（Campbell and Mankiw，1990）讨论了，为什么即使 PIH 成立，根据可利用的季度数据在误差中也能发现正的序列相关；其中有些问题可扩展到年度数据。

在时间序列应用中，利用趋势变量或 I（1）变量的增长率做 SEM 相当普遍。比如，谢伊（Shea，1993）估计了用增长率设定的产业供给曲线。

如果一个结构模型包含了时间趋势（刻画了未直接模型化的外生趋势因素），那么趋势就起到作为自己的 IV 的作用。

问题 16.4 假设你有某城市在人均鱼消费量、人均收入、鱼肉价格及鸡肉和牛肉价格等方面的月数据；收入与鸡肉和牛肉价格是外生的。假定在鱼的需求函数中不存在季节

性，但在鱼的供给中则存在季节性。你如何利用这些信息去估计一个常弹性的鱼肉需求方程？设定一个方程并讨论其识别问题。[提示：你应有鱼肉价格的11个工具变量。]

16.6 利用综列数据的联立方程模型

520 联立方程模型也会出现在综列数据的背景中。比如，我们可以像在例16.3中那样，对一群工作一段时间的人想像估计劳动供给和工资报价方程。除了容许每个时期内变量的同时决定外，我们还容许每个方程中有观测不到的影响。在一个劳动供给方程中，容许存在不随时间变化而又观测不到的闲暇偏好是有帮助的。

利用综列数据估计SEM的基本方法有两步：(1) 利用固定效应变换或一阶差分消除所关心的方程中观测不到的影响；(2) 寻找变换后的方程中内生变量的工具变量。这可能极具挑战性，因为为使分析具有说服力，我们需要找到随时间而变化的工具。为探究其原因，写下一个用于综列数据的SEM：

$$y_{it1} = \alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\beta}_1 + a_{i1} + u_{it1} \tag{16.37}$$

$$y_{it2} = \alpha_2 y_{it1} + \mathbf{z}_{it2}\boldsymbol{\beta}_2 + a_{i2} + u_{it2} \tag{16.38}$$

式中，i 为横截面；t 为时间期间；$\mathbf{z}_{it1}\boldsymbol{\beta}_1$ 或 $\mathbf{z}_{it2}\boldsymbol{\beta}_2$ 则为每个方程中一系列外生解释变量的线性函数。最一般的分析容许观测不到的影响 a_{i1}和 a_{i2}与所有解释变量甚至 $\mathbf{z}$ 中的元素都相关。我们假定特有的结构误差 u_{it1}和 u_{it2}与两个方程和所有时期的 $\mathbf{z}$ 都不相关；这就是说 $\mathbf{z}$ 是外生的。除非在特别的环境下，否则 y_{it2}与 u_{it1}相关，而 y_{it1}与 u_{it2}相关。

假设我们感兴趣的是方程(16.37)。由于合成误差 $a_{i1}+u_{it1}$潜在地与所有解释变量都相关，所以我们不能用OLS估计它。假设我们对时间差分以去掉不可观测的影响 a_{i1}：

$$\Delta y_{it1} = \alpha_1 \Delta y_{it2} + \Delta \mathbf{z}_{it1}\boldsymbol{\beta}_1 + \Delta u_{it1} \tag{16.39}$$

(和通常差分或去时间均值一样，我们只能估计至少对某些横截面单元来说是随时间而变化的变量的影响。) 现在，根据假定，这个方程中的误差项与 $\Delta \mathbf{z}_{it1}$无关，但 Δy_{it2}和 Δu_{it1}很可能相关。因此，我们需要 Δy_{it2}的一个工具变量。

就像纯横截面数据或时间序列数据的情形一样，可能的工具变量来自另一个方程：在 $\mathbf{z}_{it2}$中而又不在 $\mathbf{z}_{it1}$中的元素。实践中，我们需要 $\mathbf{z}_{it2}$中随时间而变化而又不在 $\mathbf{z}_{it1}$中的元素。这是因为我们需要 Δy_{it2}的一个工具变量，而一个变量从一个时期到另一个时期的变化不太可能与外生变量的水平值高度相关。实际上，如果将式(16.38)差分，那我们看到 Δy_{it2}自然而然的工具

变量是那些在 $\Delta \mathbf{z}_{it2}$ 中而又不在 $\Delta \mathbf{z}_{it1}$ 中的元素。

作为可能出现这种问题的一个例子，考虑例 16.3 中劳动供给函数的一个综列数据形式。假设我们在差分后得到方程

$$\Delta hours_{it} = \beta_0 + \alpha_1 \Delta \log(wage_{it}) + \Delta(\text{其他因素}_{it})$$

521 而我们希望用 $\Delta exper_{it}$ 作为（$\log \Delta wage_{it}$）的一个工具变量。问题是，由于我们在考察每个时期都工作的人，所以对所有的 i 和 t 都有 $\Delta exper_{it}=1$。（每个人在一年过去后，就都多了一年工作经历。）我们不能用一个对所有的 i 和 t 都取相同值的 IV，因此我们必须另找一个。

通常，对一项实验项目的参与可用于获得综列数据情况下的 IV。在例 15.10 中，我们用在职培训津贴的获得作为决定在职培训对工人生产力影响中培训小时数变化的一个工具变量。实际上，我们在做 SEM 的考虑下可以认为，虽然在职培训和工人生产力是联合决定的，但得到在职培训津贴在方程（15.57）中却是外生的。

在综列数据的应用研究中，我们有时也能想出一个聪明的、令人信服的工具变量，如下例子可作说明。

例 16.8　关押人数对暴力犯罪率的影响

为了估计关押人数的增加对州一级犯罪率的影响，莱维特（Levitt, 1996）使用监狱过度拥挤的诉讼案件数作为关押人数增加的工具变量。莱维特估计的方程是一阶差分的形式；我们因此可以写出其背后的固定效应模型为

$$\log(crime_{it}) = \theta_t + \alpha_1 \log(prison_{it}) + \mathbf{z}_{it1}\boldsymbol{\beta}_1 + a_{i1} + u_{it1} \tag{16.40}$$

式中，θ_t 为不同时间的截距；而 *crime* 和 *prison* 则以十万人为单位度量。（关押人数变量在上一年的最后一天度量。）向量 $\mathbf{z}_{it1}$ 包含了莱维特的文章中列出的其他控制变量，包括对人均警察数、人均收入、失业率、种族及城市和年龄分布比例的度量。

将式（16.40）差分即得到莱维特所估计的方程：

$$\Delta\log(crime_{it}) = \xi_t + \alpha_1 \Delta\log(prison_{it}) + \Delta\mathbf{z}_{it1}\boldsymbol{\beta}_1 + \Delta u_{it1} \tag{16.41}$$

犯罪率和关押人数（更准确地讲是增长率）的联立性，使得对式（16.41）的 OLS 估计通常都是不一致的。利用暴力犯罪率和莱维特所用数据（在文件 PRISON.RAW 中，1980—1993 年共 51·14=714 个观测）的一个子集，我们得到 α_1 的混合 OLS 估计值 -0.179（se=0.048）。我们还用混合 2SLS 估计了式（16.41），其中 $\Delta\log(prison)$ 的工具变量是两个二值变量，分别表示在本年或过去两年是否对过度拥挤诉讼作出最后判决。α_1 的混合 2SLS 估计值为 -1.020（se=0.366）。因此，2SLS 估计的影响要大得多；无足为奇，它也不准确得多。莱维特（Levitt, 1996）利用更长期的数据（但缺失某些州早期的数据）和更多的工具变量，也发现了类似的结论。

522 用综列数据估计 SEM 的另一种方法是，使用固定效应变换，然后使用一个像 2SLS 那样的 IV 方法。类似的程序是用 2SLS 估计具有如下形式的去时间均值方程

$$\ddot{y}_{it1}=\alpha_1\ddot{y}_{it2}+\mathbf{z}_{it1}\boldsymbol{\beta}_1+u_{it1},\quad t=1,2,\cdots,T \tag{16.42}$$

式中，$\mathbf{z}_{it1}$和$\mathbf{z}_{it1}$[*]都是工具变量。这就等价于在一个虚拟变量表达式中使用 2SLS，用单元特有的虚拟变量充当其自身的工具。艾尔斯和莱维特（Ayres and Levitt，1998）用 2SLS 在一个去时间均值方程中估计了 Lojack 电子防盗装置对城市汽车失窃率的影响。如果直接估计式（16.42），那么所需要的 df 要修正为 $N(T-1)-k$，其中 k_1 是 α_1 和 $\boldsymbol{\beta}_1$ 中元素的总个数。把单元特有的虚拟变量包括进来并对原始数据应用混合 2SLS，即产生正确的 df。

▶ 小 结

当系统中每个方程都具有其他条件不变的解释时，使用联立方程模型就是适当的。个别方程分别描述了一个市场的不同方面或不同经济个体之间的关系的情况都是很好的例子。虽然供给与需求方程是主要例子，但 SEM 在经济学和社会科学中还有许多其他的应用。

SEM 的一个重要特征是，在对系统进行充分说明后，哪个变量被假定为外生的和每个方程中出现哪些变量都很清楚。给定一个完全的系统，我们可以决定哪些方程可识别（即能被估计）。在两方程系统的重要情形中，很容易说明对第一个方程的识别条件：第一个方程至少排除了一个在第二个方程中系数非零的外生变量。

就像我们从前面的章节中知道的那样，对一个包含内生解释变量的方程的 OLS 估计，一般都会给出有偏误和不一致的估计量。相反，2SLS 则可用于估计系统中任何一个可识别的方程。有一些更深入的系统方法可供使用，但超出了我们分析的范围。

遗漏变量和联立性在应用中的区别并不总是那么鲜明。且不说测量误差，这两个问题都可以出现在同一方程中。一个很好的例子就是已婚妇女的劳动供给。受教育年数（*educ*）同时出现在劳动供给和工资报价函数中［参见方程（16.20）和方程（16.21）］。如果在劳动供给函数的误差项中有遗漏了的能力变量，那么工资和受教育程度都是内生的。重要的一点是，一个方程也能用 2SLS 去估计。

SEM 也可以用于时间序列数据。像 OLS 估计一样，我们在应用 2SLS 时必须警惕趋势性和自积过程。诸如序列相关的问题可以像在 15.7 节中那样处理。我们还给出了一个如何利用综列数据估计一个 SEM 的例子，其中

---

[*] 公式中没有 z_{it2}，恐怕是原书中的一个错误。——译者注

523 的方程为了消除不可观测的影响而做了一阶差分。然后我们就像在第 15 章中那样用混合 2SLS 估计了差分方程。另外，在某些情况下，我们可以将包括 IV 在内所有变量去时间均值，然后用混合 2SLS；这等价于对每个横截面观测引进一个虚拟变量后使用 2SLS，其中的虚拟变量作为其自身的工具变量。SEM 在综列数据中的应用是强有力的，因为它们允许我们在处理联立性的同时又控制观测不到的异质性。所以这种应用越来越常见，而且也不是特别难以估计的。

关键术语

内生变量
排除性约束
外生变量
可识别方程
恰好识别方程
滞后内生变量
阶条件
过度识别方程
前定变量
秩条件
约简（诱导）型
约简型误差
约简型参数
联立性
联立性偏误
联立方程模型（SEM）
结构方程
结构误差
结构参数
不可识别方程

习 题

16.1 写出一个“供给与需求形式”的两方程系统，即方程的左边都是变量 y_1（具体地讲是“数量”）：

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$
$$y_1 = \alpha_2 y_2 - \beta_2 z_2 + u_2$$

(i) 若 $\alpha_1 = 0$ 或 $\alpha_2 = 0$，解释为什么存在 y_1 的一个约简型。（记住 y_1 的一个约简型表达式就是外生变量和结构误差的一个线性函数。）若 $\alpha_1 \neq 0$ 和 $\alpha_2 \neq 0$，求出 y_2 的约简型。

(ii) 若 $\alpha_1 \neq 0$，$\alpha_2 \neq 0$ 和 $\alpha_1 \neq \alpha_2$，求出 y_1 的约简型。在这种情形下，y_2 有约简型吗？

(iii) 在供给与需求的例子中，$\alpha_1 \neq \alpha_2$ 的条件有可能满足吗？请解释。

16.2 令 *corn* 表示县一级人均消费玉米的蒲式耳数，*price* 表示每蒲式耳玉米的价格，*income* 表示全县人均收入，而令 *rainfall* 表示在玉米成长最

后季节的降雨英寸数。如下联立方程模型施加了供需相等的条件：

$$corn = \alpha_1 price + \beta_1 income + u_1$$
$$corn = \alpha_2 price + \beta_2 rainfall + u_2$$

哪个是供给方程，哪个是需求方程？请解释。

16.3 在第 3 章的习题 3.3 中，为对一个个人的随机样本检验每周花在 524 睡眠上的分钟数（*sleep*）和每周花在工作上的分钟数（*totwrk*）之间的交替关系，我们估计了一个方程。方程中还包括受教育水平和年龄。由于 *sleep* 和 *totwrk* 是每个人同时选择的，所估计的睡眠和工作之间的交替关系会遭到“联立性偏误”的批评吗？请解释。

16.4 假设年收入和酒消费量由 SEM

$$\log(earnings) = \beta_0 + \beta_1 alcohol + \beta_2 educ + u_1$$
$$alcohol = \gamma_0 + \gamma_1 \log(earnings) + \gamma_2 educ + \gamma_3 \log(price) + u_2$$

决定。式中，*price* 为酒的地方价格指数（包括州和地方税收）。假定 *educ* 和 *price* 都是外生的。如果 β_1、β_2、γ_1、γ_2 和 γ_3 都异于零，那么哪个方程可以识别？你如何估计这个方程。

16.5 一个决定避孕套使用在降低有性行为的高中生之间传播性病的有效性的简单模型为

$$infrate = \beta_0 + \beta_1 conuse + \beta_2 percmale + \beta_3 avginc + \beta_4 city + u_1$$

式中，*infrate* 为有性行为的学生中感染性病的比例；*conuse* 为声称合理地使用了避孕套的男孩子比例；*avginc* 为平均家庭收入；*city* 为一个表示所在学校是否处在城里的虚拟变量。这个模型是在学校这个层次上做的（即以学校为观测单元——译者注）。

(i) 在因果性和其他条件不变的模式下解释上述方程，β_1 的符号应该是什么？

(ii) 为什么 *infrate* 和 *conuse* 可能是联合决定的？

(iii) 如果避孕套使用率随着性病感染率的提高而提高，所以在方程

$$conuse = \gamma_0 + \gamma_1 infrate + \text{其他因素}，\ \gamma_1 > 0$$

中，用 OLS 估计 β_1 时可能的偏误是什么？

(iv) 令 *condis* 表示一个二值变量，若学校有分发避孕套项目则取值 1。解释这如何通过 IV 用于估计 β_1（和其他系数）。我们必须在每个方程中对 *condis* 做怎样的假定？

16.6 考虑一个雇主是否根据工人参加工会的百分比及其他因素而提供养老金计划的线性概率模型：

$$pension = \beta_0 + \beta_1 percunion + \beta_2 avgage + \beta_3 avgeduc + \beta_4 percmale + \beta_5 percmarr + u_1$$

(i) 为什么 *percunion* 可能与 *pension* 联合决定？

(ii) 假设你可以对企业工人进行调查，并搜集工人家庭的信息。你能否

想出可用于构造 *percunion* 的一个 IV 的信息？

（iii）你如何检验你想到的变量是否至少是 *percunion* 的一个合理的 IV？

525 16.7 一所大学要求你估计对女子篮球比赛门票的需求。你能搜集到 10 个赛季总共约 150 次观测的时间序列数据。一个可能的模型是

$$lATTEND_t = \beta_0 + \beta_1 lPRICE_t + \beta_2 WINPERC_t + \beta_3 RIVAL_t + \beta_4 WEEKEND_t + \beta_5 t + u_t$$

式中，$PRICE_t$ 为门票价格（可能以真实价格度量，比如通过地区消费价格指数进行平缩）；$WINPERC_t$ 为球队当前获胜的概率；$RIVAL_t$ 为一个标志着比赛是否势均力敌的虚拟变量；$WEEKEND_t$ 为一个标志着球赛是否在周末进行的虚拟变量。l 表示自然对数，所以这个需求函数具有常价格弹性。

（i）为什么在这个方程中有一个时间趋势是个好想法？

（ii）门票供给由体育馆的容量所决定；假定这个供给 10 年不变，这意味着供给的数量不随价格而变化。这意味着价格在这个需求方程中必然是外生变量吗？［提示：回答是否定的。］

（iii）假设门票的名义价格缓慢变化（如在每个赛季之初）。体育委员会部分基于上赛季的平均售票和该队上赛季的胜率来选择价格。在什么样的条件下，上赛季的胜率（$SEASPERC_{t-1}$）是 $lPRICE$ 的一个有效的工具变量？

（iv）在方程中包括男子篮球比赛的真实价格（的对数）看起来合理吗？请解释。经济理论预测其系数的符号是什么样的？你能想到另外一个与男子篮球相关而又属于女子观众方程的变量吗？

（v）如果你担心某些序列（特别是 $lATTEND$ 和 $lPRICE$）有单位根，你如何改变所估计的方程？

（vi）如果某些比赛的门票售完，这会导致估计需求方程出现什么问题？［提示：如果门票售完，你一定观察到真实需求了吗？］

16.8 平均每个学生的学校支出对当地住房价值的影响有多大？令 $HPRICE$ 表示校区内住房价格的中位数，而令 $EXPEND$ 表示平均每个学生的支出。利用 1992 年、1994 年和 1996 年的综列数据，我们提出模型

$$lHPRICE_{it} = \theta_t + \beta_1 lEXPEND_{it} + \beta_2 lPOLICE_{it} + \beta_3 lMEDINC_{it} + \beta_4 PROPTAX_{it} + a_{i1} + u_{it1}$$

式中，$POLICE_{it}$ 为人均警力支出；$MEDINC_{it}$ 为收入中位数；$PROPTAX_{it}$ 为财产税率；l 为自然对数。由于住房价值直接影响学校可用于各项开支的收益，所以支出和住房价格同时决定。

假令学校的融资方式在 1994 年发生巨大变化：不是征收地方财产税，学校的资金来源基本上由州一级单位决定，令 $lSTATEALL_{it}$ 表示州政府在第 t 年对地区 i 津贴的对数，一旦我们控制了支出和地区的固定影响，上述方程中哪个是外生变量？你如何估计 β_1？

计算机习题

526 16.9 本题用到 SMOKE.RAW 中的数据。

(i) 估计吸烟对年收入的影响（可能通过因病损失的工作日或生产力影响）的一个模型是

$$\log(income)=\beta_0+\beta_1 cigs+\beta_2 educ+\beta_3 age+\beta_4 age2+u_1$$

式中，*cigs* 为平均每天吸烟的数量。你如何解释 β_1？

(ii) 为了反映香烟消费可能与收入同时决定，一个香烟需求方程是

$$cigs=\gamma_0+\gamma_1\log(income)+\gamma_2 educ+\gamma_3 age+\gamma_4 age^2+\gamma_5\log(cigpric)+\gamma_6 restaurn+u_2$$

式中，*cigpric* 为每包香烟的价格（美分）；*restaurn* 为一个二值变量，并在这个人所定居的州有餐馆吸烟限制时等于 1。假定这些变量对个人而言都是外生的，你预期 γ_5 和 γ_6 具有什么样的符号？

(iii) 在什么样的条件下第 (i) 部分的收入方程可识别？

(iv) 用 OLS 估计收入方程并讨论 β_1 的估计值。

(v) 估计 *cigs* 的约简型。（记住这就要求将 *cigs* 对所有内生变量回归。）log（*cigpric*）和 *restaurn* 在约简型中显著吗？

(vi) 现在用 2SLS 估计收入方程。讨论 β_1 的估计值与 OLS 估计值的比较。

(vii) 你认为香烟价格和餐馆吸烟限制在收入方程中是外生的吗？

16.10　本题用到 MROZ.RAW 中的数据。

(i) 用 $\log(hours)$ 作为因变量重新估计例 16.5 中的劳动供给函数。将估计出的弹性（现在是常数）与方程（16.24）在平均工作小时数处所得到的估计值相比较。

(ii) 在第 (i) 部分的劳动供给方程中，容许 *educ* 因遗漏了能力变量而成为外生变量。用 *motheduc* 和 *fatheduc* 作为 *educ* 的 IV。记住，你现在在方程中有两个内生变量。

(iii) 检验第 (ii) 部分 2SLS 估计中过度识别约束。这些 IV 通过了检验吗？

16.11　本题用到 OPENNESS.RAW 中的数据。

(i) 由于 $\log(pcinc)$ 在方程（16.22）和 *open* 的约简型中都是不显著的，所以将它从分析中去掉。用 OLS 和 IV 在没有 $\log(pcinc)$ 的情况下估计方程（16.22）。重要的结论有什么变化吗？

527 (ii) 仍将 $\log(pcinc)$ 放在分析之外，*land* 或 $\log(land)$ 是 *open* 更好的工具变量吗？[提示：将 *open* 对二者分别并同时回归。]

(iii) 现在回到方程（16.22）中，在方程中增加虚拟变量 *oil* 并视之为外生变量，用 IV 估计方程。作为一个石油生产国对通货膨胀有其他条件不变的影响吗？

16.12　本题用到 CONSUMP.RAW 中的数据。

(i) 在例 16.7 中，用 15.5 节的方法检验在估计方程（16.35）时的那一个过度识别约束。你的结论是什么？

(ii) 由于潜在的数据度量问题和信息滞后，坎贝尔和曼基（Campbell

and Mankiw, 1990）使用所有变量的二阶滞后值作为工具变量。只用 gc_{t-2}，gy_{t-2}和 $r3_{t-2}$作为工具变量重新估计方程（16.35）。这些估计值与方程（16.36）中的那些估计值相比如何？

（iii）将 gy_t 对第（ii）部分的 IV 回归并检验 gy_t 与它们是否充分相关。这一点为什么重要？

16.13　利用总统经济报告（1998 年或以后）更新 CONSUMP.RAW 中的数据，至少截止到 1996 年。重新估计方程（16.35）。重要的结论有什么变化吗？

16.14　本题用到 CEMENT.RAW 中的数据。

（i）水泥价格月增长率（*gprc*）作为供给数量增长率（*gcem*）函数的一个静态（逆）供给函数是

$$gprc_t = \alpha_1 gcem_t + \beta_0 + \beta_1 grprcpet + \beta_2 feb_t + \cdots + \beta_{12} dec_t + u_t^s$$

式中，*grprcpet*（汽油价格上涨率）被假定为外生变量，而 *feb*，…，*dec* 为月度虚拟变量。你预期 α_1 和 β_1 的符号是什么？用 OLS 估计这个方程。供给函数右上倾斜吗？

（ii）变量是美国真实国防支出的月增长率。*grdefs* 要作为 *gcem* 的一个好的工具变量，你需要对它做什么假定？检验 *gcem* 是否与 *grdefs* 偏相关。（不用担心约简型中可能的序列相关。）你能用 *grdefs* 作为估计供给函数中的一个 IV 吗？

（iii）谢伊（Shea，1993）认为建住宅房的产出增长率（*grres*）和非住宅房的产出增长率（*grnon*）是 *gcem* 的有效工具变量。其思想是，存在一些应该与供给误差 u_t^s 大致无关的需求移动因子。检验 *gcem* 是否与 *grres* 和 *grnon* 偏相关；同样不用担心约简型中的序列相关。

（iv）利用 *grres* 和 *grnon* 作为 *gcem* 的工具变量估计供给函数。你对水泥的静态供给函数得到什么结论？［动态供给函数显然是右上倾斜的；参见 Shea（1993）。］

16.15　参考例 13.9 并利用 CRIME4.RAW 中的数据。

528 （i）假定你在做差分以消除不可观测的影响之后，认为 Δlog（*polpc*）与 Δlog（*crmrte*）是同时决定的；特别是犯罪的增加与警察人数的增加有关系。这对解释方程（13.33）中 Δlog（*polpc*）的正系数有何帮助？

（ii）变量 *taxpc* 表示全县人均征税量。将它排除在犯罪方程之外看上去合理吗？

（iii）包括潜在的工具变量 Δlog（*taxpc*）后利用混合 OLS 估计 Δlog（*polpc*）的约简型。Δlog（*taxpc*）看起来是很好的一个候选 IV 吗？

（iv）假设在几年后，北卡罗来纳州对某些县津贴以扩大其警察规模。你如何利用这个信息估计增加的警察对犯罪率的影响？

第 17 章　限值因变量模型和样本选择纠正

529　我们在第 7 章研究了线性概率模型，它不过是多元回归模型在二值因变量情况下的一个应用。二值因变量只是**限值因变量**(limited dependent variable, LDV)之一例。广义而言，一个 LDV 就是一个取值范围明显受到限制的因变量。二值变量只取 0 和 1 两个值，当然是 LDV。我们还看到过其他限值因变量的例子：养老金参与率必须介于 0～100 之间；一个人在给定年份被拘捕的次数也是一个非负的整数；而大多数大学的大学 *GPA* 也介于 0～4.0 之间。

我们欲解释的大多数经济变量都以某种方式受到限制，通常都要求它们必须为正值。例如，小时工资、住房价格和名义利率都必须大于零，但并非所有这种变量都需要特别处理。如果一个严格为正的变量取许多不同的值，那么很少需要特殊的计量模型。

而当 y 是一个离散而又取少数几个值时，把它看成一个近似连续的变量就毫无意义。y 的离散性本身并不意味着线性模型就不合适。但如我们在第 7 章对二值响应的讨论中所见，线性概率模型有些缺陷。在 17.1 节，我们讨论 logit 和 probit 模型，它们都克服了 LPM 的缺陷；但其不足之处在于更难以解释。

计量经济分析中还会出现其他类型的限值因变量，特别是在建立个人、家庭和企业行为的模型时。优化行为常常会导致总体中不可忽略的一部分的**角点解**(corner solutions)；也就是说，选择数量零或零美元是最优的。比如，

在任一给定年份，有相当数量的家庭的慈善捐助为零。因此，虽然年度家庭慈善捐助的总体分布散布于一个很大的正数范围内，但在数字零上却相当集中。尽管线性模型可能适合于刻画慈善捐助的期望值，但线性模型又可能对某些家庭作出负值的预测。由于许多观测都是零，所以就不可能取对数。我们在17.2节讨论的托比模型，显然就是为模型化角点解因变量而设计的。

另一类重要的LDV是只取值非负整数的计数变量。17.3节将说明，泊松回归模型如何很好地模型化计数变量。

530 在某些情形下，我们观测到限值因变量是由于对数据进行了截取，这个问题我们将在17.4节介绍。在17.5节探讨的样本选择一般性问题中，我们观察到的是潜在总体的一个非随机样本。

虽然限值因变量模型也能用于时间序列和综列数据中，但通常它们都用于横截面数据。样本选择问题通常都源于横截面或综列数据。在本章，我们只关注其在横截面数据中的应用。伍德里奇（Wooldridge, 1999）则在综列数据模型的背景下提出了这些问题，并对在横截面数据和综列数据中的应用进行了更详尽的探讨。

17.1 二值响应的logit和probit模型

虽然估计和使用线性概率模型很简单，但它有我们在7.5节所讨论的那些缺陷。最重要的两个不足是，拟合出来的概率可能小于0或大于1，任何一个解释变量（以水平值形式出现）的偏效应都是不变的。使用更复杂的**二值响应模型**（binary response models）可以克服LPM的这些缺陷。

在一个二值响应模型中，我们所关注的核心基本上是响应概率

$$P(y=1|\mathbf{x}) = P(y=1|x_1, x_2, \cdots, x_k) \qquad (17.1)$$

式中，我们用$\mathbf{x}$表示全部解释变量所构成的集合。比如，当y是一个就业指标时，$\mathbf{x}$可能就包括诸如受教育程度、年龄、婚姻状况（包括是否参加了最近一次工作培训项目的一个二值标示变量）和其他影响就业状况的各种个人特征。

设定logit和probit模型

在LPM中，我们假定响应概率对一系列参数β_j是线性的，见方程（7.27）。为避免LPM的局限性，考虑形如

$$P(y=1|\mathbf{x}) = G(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k) = G(\beta_0+\mathbf{x}\boldsymbol{\beta}) \qquad (17.2)$$

的一类二值响应模型，其中G是一个取值范围严格介于0～1之间的函数：对所有实数z，都有$0<G(z)<1$。这就确保估计出来的响应概率严格地介

于 0～1 之间。如在前面的章节中一样，这里 $x\boldsymbol{\beta} = \beta_1 x_1 + \cdots + \beta_k x_k$。

为了保证概率介于 0～1 之间，已经提出了函数 G 的各种非线性形式。我们这里将讨论的两个函数形式，它们（与 LPM 一起）被使用于绝大多数应用之中。在 **logit 模型**（logit model）中，G 是对数函数： 531

$$G(z) = \exp(z)/[1 + \exp(z)] = \Lambda(z) \tag{17.3}$$

对所有的实数 z，它都介于 0～1 之间。它是一个标准的逻辑斯蒂随机变量的累积分布函数。在 **probit 模型**（probit model）中，G 是标准正态的累积分布函数（cdf），可表示为积分

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)\mathrm{d}v \tag{17.4}$$

式中，$\phi(z)$ 为标准正态密度函数：

$$\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2) \tag{17.5}$$

G 的这个选择也确保了式(17.2)对所有参数和 x_j 的值都严格介于 0～1 之间。

式(17.3)和式(17.4)中的 G 函数都是增函数。它们都是在 $z = 0$ 时增加得最快；在 $z \to -\infty$ 时，$G(z) \to 0$；而在 $z \to \infty$ 时，$G(z) \to 1$。逻辑斯蒂函数画在图 17.1 中，而标准正态 cdf 的形状与逻辑斯蒂的 cdf 十分相似。

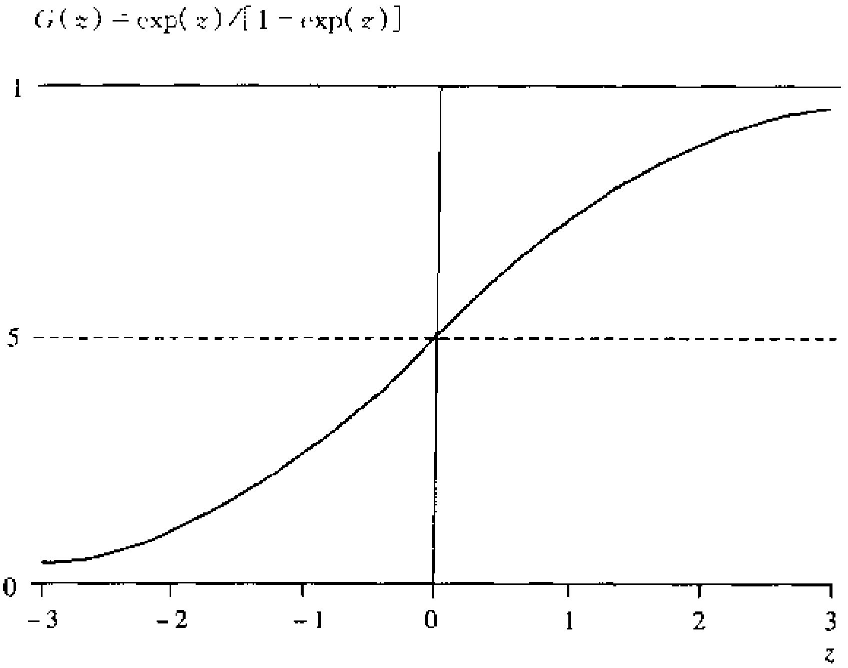


图 17.1　逻辑斯蒂函数 $G(z) = \exp(z)/[1 + \exp(z)]$ 的图示

532 logit 和 probit 模型都可以从一个满足经典线性模型假定的**潜变量模型**（latent variable model）推导出来。令 y^* 为一个由

$$y^* = \beta_0 + x\boldsymbol{\beta} + e, \quad y = 1[y^* > 0] \tag{17.6}$$

决定的观测不到的变量或潜变量。我们在其中引入记号 1［·］来定义一个二值结果。函数 1［·］被称为标示函数，括号中的事件正确时取值 1，在其他情况下取值 0。因此，若 $y^*>0$，则 y 为 1；若 $y^*\leqslant 0$，则 y 为 0。我们假定 e 独立于 $\boldsymbol{x}$、并服从标准的逻辑斯蒂分布或标准正态分布。但无论在哪种情况下，e 都对称于 0 而分布，这意味着对所有的实数 z 都有 $1-G(-z)=G(z)$。经济学家倾向于 e 的正态假定，这就是计量经济学中 probit 模型比 logit 模型更为普遍使用的原因。此外，由于正态分布的性质，用 probit 模型就更容易分析我们以后将讨论的几个设定问题。

从式（17.6）和给定的假定，我们可以推导出 y 的响应概率：

$$\begin{aligned}\mathrm{P}(y=1\mid\boldsymbol{x})&=\mathrm{P}(y^*>0\mid\boldsymbol{x})=\mathrm{P}[e>-(\beta_0+\boldsymbol{x\beta})\mid\boldsymbol{x}]\\&=1-\mathrm{G}[-(\beta_0+\boldsymbol{x\beta})]=\mathrm{G}(\beta_0+\boldsymbol{x\beta})\end{aligned}$$

恰好与式(17.2)相同。

在二值响应模型的大多数应用中，主要目的都是为了解释 x_j 对响应概率 $\mathrm{P}(y=1\mid\boldsymbol{x})$ 的影响。而潜变量表达式给我们留下的印象是，我们主要关心每个 x_j 对 y^* 的影响。如我们将见到的，对 logit 和 probit 模型而言，x_j 对 $\mathrm{E}(y^*\mid\boldsymbol{x})=\beta_0+\boldsymbol{x\beta}$ 和 $\mathrm{E}(y\mid\boldsymbol{x})=\mathrm{P}(y=1\mid\boldsymbol{x})=\mathrm{G}(\beta_0+\boldsymbol{x\beta})$ 影响的方向总是一致的。但潜变量 y^* 很难具有一个良好定义的度量单位。（比如，y^* 可能是从两个不同的行动中所得到的效用水平之差。）因此，每个 β_j 的大小本身并非特别有用（与线性概率模型形成鲜明对照）。多数情况下，我们想估计 x_j 对成功概率 $\mathrm{P}(y=1\mid\boldsymbol{x})$ 的影响，但由于 $G(\cdot)$ 的非线性，所以这个问题有些复杂。

要得到基本上连续的变量对响应概率的偏效应，我们就必须用微积分。如果 x_j 是一个大致连续的变量，那它对 $p(\boldsymbol{x})=\mathrm{P}(y=1\mid\boldsymbol{x})$ 的偏效应可通过如下偏导数得到：

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j}=g(\beta_0+\boldsymbol{x\beta})\beta_j,\quad g(z)\equiv\frac{\mathrm{d}G}{\mathrm{d}z}(z)\tag{17.7}$$

由于 G 是一个连续随机变量的 cdf，所以 g 是一个概率密度函数。在 logit 和 probit 的情形中，$G(\cdot)$ 是一个严格递增的 cdf，于是对所有的 z 都有 $g(z)>0$。因此，x_j 对 $p(\boldsymbol{x})$ 的偏效应通过正量 $g(\beta_0+\boldsymbol{x\beta})$ 而取决于 $\boldsymbol{x}$，这就意味着偏效应总具有与 β_j 一样的符号。

方程(17.7)表明，任何两个连续解释变量的相对影响都与 $\boldsymbol{x}$ 无关：x_j 和 533 x_h 的偏效应之比为 β_j/β_h。在 g 关于 0 对称分布（惟一的众数是 0）的典型情形中，在 $\beta_0+\boldsymbol{x\beta}=0$ 时出现最大的影响。例如，在 $g(z)=\phi(z)$ 的 probit 情形中，$g(0)=\phi(0)=1/\sqrt{2\pi}\approx 0.40$。在 logit 情形中，$g(z)=\exp(z)/[1+\exp(z)]^2$，所以 $g(0)=0.25$。

比方说，如果 x_1 是一个二值解释变量，那么在保持其他变量不变的情况下，x_1 从 0 变化到 1 的偏效应无非是

$$G(\beta_0+\beta_1+\beta_2x_2+\cdots+\beta_kx_k)-G(\beta_0+\beta_2x_2+\cdots+\beta_kx_k)\tag{17.8}$$

这里同样取决于所有其他变量 x_j 的值。比如，若 y 是一个就业标示，而 x_1

是表示是否参加工作培训项目的一个二值虚拟变量，那么式(17.8)就反映了就业概率因工作培训项目而发生的改变；这还取决于受教育程度和工作经历等其他影响就业能力的变量。注意到，知道了 β_1 的符号，就足以决定此项目的影响是正还是负。但为了得到影响的大小，我们还必须估计式(17.8)中的数量。

我们也可以对其他离散变量(如子女数量)用式(17.8)中的差分。若 x_k 表示这个变量，则 x_k 从 c_k 变化到 c_k+1 对概率的影响就是

$$\begin{aligned}&G[\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k(c_k+1)]\\&\quad -G(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k c_k)\end{aligned}\tag{17.9}$$

很容易在方程中把标准的函数形式包括进解释变量之中。比如，在模型

$$\mathrm{P}(y=1|\boldsymbol{z})=G[\beta_0+\beta_1 z_1+\beta_2 z_1^2+\beta_3\log(z_2)+\beta_4 z_3]$$

中，z_1 对 $\mathrm{P}(y=1|\mathrm{z})$ 的偏效应是 $\partial\mathrm{P}(y=1|z)/\partial z_1=\mathrm{g}(\beta_0+\boldsymbol{x\beta})(\beta_1+2\beta_2 z_1)$，$z_2$ 对响应概率的偏效应是 $\partial\mathrm{P}(y=1|z)/\partial z_2=\mathrm{g}(\beta_0+\boldsymbol{x\beta})(\beta_3/z_2)$，其中 $\boldsymbol{x\beta}=\beta_1 z_1+\beta_2 z_1^2+\beta_3\log(z_2)+\beta_4 z_3$。对含有解释变量交互项(包括离散变量和连续变量之间的交互项)的模型都可类似处理。在度量离散变量的影响时，我们应该使用式(17.9)。

logit 和 probit 模型的最大似然估计

我们该如何估计非线性的二值响应模型呢？要估计 LPM，我们可以使用普通最小二乘法(参见 7.5 节)或加权最小二乘法(参见 8.5 节)。由于 $\mathrm{E}(y|\boldsymbol{x})$ 的非线性性质，所以 OLS 和 WLS 都不适用。虽然我们可以使用这些方法的非线性形式，但使用**最大似然估计**(maximum likelihood estimation，MLE)也不会更复杂(对最大似然估计的简单讨论可参见附录 B)。到目前为止，我们还不怎么需要 MLE，尽管我们确实注意到，在经典线性模型假定下，OLS 估计量也是最大似然估计量(以解释变量为条件)。但在估计受限因变量模型时，最大似然估计的方法则必不可少。

534 假定有一个容量为 n 的随机样本。为了得到以解释变量为条件的最大似然估计量，我们需要 y_i 在给定 $\boldsymbol{x}_i$ 下的密度函数。可以把它写成

$$f(y|\boldsymbol{x}_i;\boldsymbol{\beta})=[G(\boldsymbol{x}_i\boldsymbol{\beta})]^y[1-G(\boldsymbol{x}_i\boldsymbol{\beta})]^{1-y},y=0,1\tag{17.10}$$

为简单起见，把截距放到向量 $\boldsymbol{x}_i$ 中。我们很容易看到，当 $y=1$ 时，得到 $G(\boldsymbol{x}_i\boldsymbol{\beta})$；而当 $y=0$ 时，得到 $1-G(\boldsymbol{x}_i\boldsymbol{\beta})$。第 i 次观测的对数似然函数是参数和数据 $(\boldsymbol{x}_i, y_i)$ 的函数，可通过对式(17.10)取对数而得到

$$l_i(\boldsymbol{\beta})=y_i\log[G(\boldsymbol{x}_i\boldsymbol{\beta})]+(1-y_i)\log[1-G(\boldsymbol{x}_i\boldsymbol{\beta})]\tag{17.11}$$

由于 $G(\cdot)$ 对 logit 和 probit 模型来说严格地介于 0～1 之间，所以 $l_i(\boldsymbol{\beta})$ 对所有 $\boldsymbol{\beta}$ 值都有良好的定义。

将式(17.11)对所有观测求和即得到样本容量为 n 的对数似然函数：$\mathcal{L}(\boldsymbol{\beta})=\sum_{i=1}^{n}\ell_i(\boldsymbol{\beta})$。$\boldsymbol{\beta}$ 的 MLE(记为 $\hat{\boldsymbol{\beta}}$)最大化了这个似然函数。若 $G(\cdot)$是标准 logit 的 cdf，则 $\hat{\boldsymbol{\beta}}$ 为 logit 估计量；若 $G(\cdot)$是标准正态的 cdf，则 $\hat{\beta}$ 为 probit 估计量。

由于最大化问题的非线性性质，我们写不出 logit 或 probit 最大似然估计值的公式。这除了带来计算上的问题外，还使得 logit 和 probit 模型的统计理论比 OLS 甚至 2SLS 都困难得多。不过，关于随机样本的(条件)MLE 的一般理论意味着，在很一般性的条件下，MLE 是一致的、渐近正态的和渐近有效的。[一般性的讨论，可参见 Wooldridge(1999，第 13 章)。] 我们在此只使用这些结论：如果我们理解了这些统计量的含义，应用 logit 和 probit 模型都相当容易。

每个 $\hat{\beta}_j$ 都带有一个(渐近的)标准误，其表达式很复杂，并在本章的附录中给出。一旦有了这些标准误(任何支持 logit 和 probit 的软件包都会与系数估计值一起报告这些标准误)，那我们就能像对 OLS，2SLS 和曾讨论的其他估计量一样，构造(渐近的)t 检验和置信区间。特别是，要检验 H_0：$\hat{\beta}_j=0$，我们在决定了单侧或双侧对立假设后，就能计算 t 统计量 $\hat{\beta}_j/\mathrm{se}(\hat{\beta}_j)$，并以通常的方式进行检验。

多重假设的检验

我们也可以对 logit 和 probit 模型中的多个约束进行检验。在多数情况下，都像 4.5 节一样对多个排除性约束进行检验。我们这里只考虑排除性约束。

有三种方法检验 logit 和 probit 模型中的排除性约束。拉格朗日乘数或得分检验就像在 5.2 节中的线性情形一样，只需要在虚拟假设下对模型进行估计；由于很少要用得分检验去检验排除性约束，所以我们在此不予讨论[至于得分检验在二值响应模型中的其他应用，可参见 Wooldridge(1999，第 15 章)]。

535 瓦尔德(Wald)检验只要求估计不受约束模型。在线性模型的情形下，瓦尔德统计量在进行简单变换后实质上就是 F 统计量；所以，没有必要单独讨论瓦尔德统计量。瓦尔德统计量的表达式在伍德里奇(Wooldridge，1999，第 15 章)的论述中给出。凡容许在估计不受约束模型后检验排除性约束的计量经济软件包都计算了这个统计量。它服从一个渐近的 χ 平方分布，其中 df 等于被检验的约束个数。

如果受约束模型和不受约束模型都很容易估计(如通常含有排除性约束的情形一样)，那么**似然比(likelihood ratio，LR)检验**就很有诱惑力。LR 检验基于线性模型中 F 检验同样的概念。F 检验度量了从模型中去掉变量时残差平方和的增加。LR 检验是基于不受约束模型和受约束模型的对数似然函数

之差。其思想是，由于 MLE 最大化了对数似然函数，所以去掉变量一般会导致一个较小（至少不会更大）的对数似然值。（这就像从一个回归中去掉变量不可能使 R-平方增加一样。）问题是，对数似然值的下降程度是否大到足以断定去掉的变量是重要的。一旦我们有了一个检验统计量和一系列的临界值，就能作出这个判断。

似然比统计量(likelihood ratio statistic)是对数似然值之差的 2 倍：

$$LR = 2(\mathcal{L}_{ur} - \mathcal{L}_r) \tag{17.12}$$

式中，$\mathcal{L}_{ur}$ 为受约束模型的对数似然值；而 $\mathcal{L}_r$ 为受约束模型的对数似然值。由于 $\mathcal{L}_{ur} \geqslant \mathcal{L}_r$，所以 LR 总是非负数，且通常严格为正。在计算 LR 统计量时，重要的是要知道，$\mathcal{L}_{ur}$ 和 $\mathcal{L}_r$ 都可以为负。这并不改变我们计算 LR 的方法；只是我们在计算中必须保留其负号。

之所以在式(17.12)中需要乘以 2，是为了使得 LR 在 H_0 下服从渐近 χ 平方分布。如果我们检验 q 个排除性约束，则 $LR \stackrel{a}{\sim} \chi^2_q$。这意味着，要在 5% 的显著性水平上检验 H_0，我们必须用 χ^2_q 分布中第 95 个百分位作为临界值。大多数软件包都能很容易地计算出 p 值。

问题 17.1 ☞

解释一个企业在给定年份是否被另一个企业接管的 probit 模型是

$$\mathrm{P}(takeover = 1 \mid \boldsymbol{x}) = \Phi(\beta_0 + \beta_1 avgprof + \beta_2 mktval + \beta_3 debtearn + \beta_4 ceoten + \beta_5 ceosal + \beta_6 ceoage)$$

式中，*takeover* 为一个二值响应变量；*avgprof* 为过去几年企业的平均利润率；*mktval* 为企业的市场价值；*debtearn* 为债务—收益比；*ceoten*，*ceosal* 和 *ceoage* 则分别为企业首席执行官的任期、平均年薪和年龄。陈述一个虚拟假设，表明在其他因素不变的情况下，与 CEO 有关的变量对接管概率没有影响。LR 或瓦尔德检验的 χ 平方分布中，df 是多少？

解释logit 和probit 模型的估计值

从实践的角度看，在有了现代计算机后，logit 或 probit 模型最困难的方面是表述和解释结论。所有做 logit 和 probit 的软件包都报告系数估计值及其
536 标准误和对数似然函数值，在应用研究中，都应该把这些报告出来。系数给出每个 x_j 对响应概率偏效应的符号，而 x_j 的统计显著性则由我们能否在一个足够小的显著性水平上拒绝 $\mathrm{H}_0: \beta_j = 0$ 来决定。

通常报告的一种度量拟合优度的方法是所谓的**正确预测百分数**(percent correctly predicted)，其计算方法如下：对每个 i，我们都计算 y_i 取值 1 的估计概率 $G(\hat{\beta}_0 + \boldsymbol{x}\hat{\boldsymbol{\beta}})$。若 $G(\hat{\beta}_0 + \boldsymbol{x}\hat{\boldsymbol{\beta}}) > 0.5$，则 y_i 的预测值为 1，而若 $G(\hat{\beta}_0 + \boldsymbol{x}\hat{\boldsymbol{\beta}}) \leqslant 0.5$，则 y_i 的预测值为 0。预测的 y_i 与实际的 y_i 相符的次数所占百分

比就是正确预测百分数。虽然这个度量方法多少有些用处，但很有可能在不用模型的情况下，也能得到相当高的正确预测百分数。比如，假使在一个容量为200的样本中，有180个观测的 $y_i=0$，而其中有150个用前述方法预测为零。即便对 $y_i=1$ 的观测没有一个能正确预测，我们仍正确预测了75%的结果。由于这样的例子，对两种结果都报告正确预测百分数是有意义的。

度量二值响应的方法，还有各种拟 **R-平方**(pseudo R-squared)度量。麦克法登(McFadden,1974)提出的度量是 $1-\mathcal{L}_{ur}/\mathcal{L}_0$，其中 $\mathcal{L}_{ur}$ 表示被估计模型的对数似然函数，而 $\mathcal{L}_0$ 则表示只有截距项的模型中的对数似然函数。它与OLS回归的 R-平方相似，因为OLS的 R-平方可写成 $1-\mathrm{SSR}_{ur}/\mathrm{SSR}_0$，其中 SSR_{ur} 是残差平方和，而 SSR_0 则是总平方和。虽然还有人提出过其他几种度量[如参见Maddala(1983，第2章)]，但拟合优度通常比不上解释变量的统计和经济显著性那么重要。

我们常常想估计 x_j 对响应概率 $\mathrm{P}(y=1|\mathbf{x})$ 的影响。若 x_j 是(大致)连续的，则对 x_j 的“较小”变化，有

$$\Delta\hat{P}(y=1|\mathbf{x})\approx[g(\hat{\beta}_0+\mathbf{x}\hat{\boldsymbol{\beta}})\hat{\beta}_j]\Delta x_j \tag{17.13}$$

由于 $g(\hat{\beta}_0+\mathbf{x}\hat{\boldsymbol{\beta}})$ 取决于 $\mathbf{x}$，所以我们必须在 $\mathbf{x}$ 有意义的地方计算它。通常都是将 x_j 的样本平均值代人，以得到 $g(\hat{\beta}_0+\bar{\mathbf{x}}\hat{\boldsymbol{\beta}})$。然后可用这个因子来调整每个 $\hat{\beta}_j$(或至少是连续变量的那些系数)，以得到 x_j 每提高一个单位的影响。如果 $\mathbf{x}$ 包含了某些解释变量的非线性函数，如自然对数函数或二次函数，那我们仍可以将样本平均值代人非线性函数或将非线性函数平均化。要得到总体平均的单位影响，用第一种办法讲得通。如果一个软件包自动地将系数乘以 $g(\hat{\beta}_0+\bar{\mathbf{x}}\hat{\boldsymbol{\beta}})$，那它就肯定平均了非线性函数，因为它不能告诉我们什么时候一个解释变量是某个基本变量的非线性函数。其差别基本上不大。

在计算 $g(\hat{\beta}_0+\mathbf{x}\hat{\boldsymbol{\beta}})$ 时，有时也用某些关键变量的最小和最大或上下四分位值，以便我们看出当 $\mathbf{x}$ 的某些元素变大或变小时，偏效应将如何变化。

方程(17.13)还说明了，如何大致地比较logit和probit斜率估计值的大小。像前面提到的一样，对probit而言，$g(0)\approx0.4$，而对logit而言，$g(0)\approx0.25$。因此，为了使logit和probit的斜率估计值可比较，我们或者将probit的估计值乘以 $0.4/0.25=1.6$，或者将logit的估计值乘以0.625。在线性概率模
537 型中，$g(0)$ 实际上就是1，所以logit的斜率估计值应该除以4才大致可与LPM的估计值相比；而probit的斜率估计值则应该除以2.5才大致可与LPM的估计值相比。一个更准确的比较是将probit的斜率估计值乘以 $\phi(\hat{\beta}_0+\bar{\mathbf{x}}\hat{\boldsymbol{\beta}})$，而将logit的斜率估计值乘以 $\exp(\hat{\beta}_0+\mathbf{x}\hat{\boldsymbol{\beta}})/[1+\exp(\hat{\beta}_0+\bar{\mathbf{x}}\hat{\boldsymbol{\beta}})]^2$，其中的系数估计值分别为probit或logit中的估计值。

如果说 x_k 是一个二值变量，可能代人 x_k 的0或1值而非 $\bar{x}_k$(样本中出现1的比例)才讲得通。代人二值变量的平均值意味着其影响并非真正对应于某特定的个体。但结论通常都类似，二者之间的选择完全是出于个人偏好。

如果 x_k 是一个离散变量，就可以估计它从 c_k 增加到 c_k+1 导致预测概

率的变化为

$$\begin{aligned} &G[\hat{\beta}_0+\hat{\beta}_1\bar{x}_1+\cdots+\hat{\beta}_{k-1}\bar{x}_{k-1}+\hat{\beta}_k(c_k+1)] \\ &\quad -G(\hat{\beta}_0+\hat{\beta}_1\bar{x}_1+\cdots+\hat{\beta}_{k-1}\bar{x}_{k-1}+\hat{\beta}_k c_k) \end{aligned} \tag{17.14}$$

特别地，当 x_k 是一个二值变量时，我们取 $c_k=0$。当然，我们必须决定对其他解释变量代入什么；我们一般都是对大致连续的变量代入其平均值。

538 **例 17.1　已婚妇女的劳动市场参与**

我们现在用 MROZ.RAW 数据，通过 logit 和 probit 来估计例 8.8(也可参见 7.5 节)中的劳动参与模型。我们也报告了例 8.8 中利用异方差稳健标准误得到的线性概率模型估计值。表 17.1 给出了结果(及括号中的标准误)。

表 17.1　劳动市场参与的 LMP，logit 和 probit 估计值

	因变量：*inlf*		
自变量	LPM (OLS)	Logit (MLE)	Probit (MLE)
nwifeinc	−0.0034 (0.0015)	−0.021 (0.008)	−0.012 (0.005)
educ	0.038 (0.007)	0.221 (0.043)	0.131 (0.025)
exper	0.039 (0.006)	0.206 (0.032)	0.123 (0.019)
$exper^2$	−0.000 60 (0.000 18)	−0.003 2 (0.001 0)	−0.001 9 (0.000 6)
age	−0.016 (0.002)	−0.088 (0.015)	−0.053 (0.008)
kidslt6	−0.262 (0.032)	1.443 (0.204)	−0.868 (0.119)
kidsge6	0.013 (0.013)	0.060 (0.075)	0.036 (0.043)
常数项	0.586 (0.151)	0.425 (0.860)	0.270 (0.509)
正确预测百分数	73.4	73.6	73.4
对数似然值	—	−401.77	−401.30
拟 *R*-平方	0.264	0.220	0.221

这三个模型的结论是一致的。每个模型所得到系数的符号都相同，而且统计显著的变量都一样。LPM 的拟 *R*-平方刚好就是通常对 OLS 报告的 *R*-

平方；对 logit 和 probit 而言，拟 R-平方是基于前述对数似然值而计算的。

如我们已强调的那样，各模型的系数大小不能直接比较。利用前面讨论过的经验法则，我们可以将 logit 估计值除以 4 和将 probit 估计值除以 2.5 后与 LPM 估计值比较，比如 *kidslt6* 的系数，换算后的 logit 估计值约为 -0.361，换算后的 probit 估计值约为 -0.347，都比 LPM 估计值大（原因稍后解释）。类似地，换算后 *educ* 的 logit 和 probit 估计值分别为 0.055 和 0.052；它们也多少比 LPM 估计值 0.038 大些，但相差不大。

如果我们在样本中自变量的平均值（包括 $exper^2$ 的平均值）处计算标准正态概率密度函数 $\phi(\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_kx_k)$，结果近似为 0.391，与 0.4 足够接近，使得换算 probit 估计值的经验法则在求对响应概率的影响时有用。换句话说，为了估计任一自变量一个单位的提高对响应概率的改变，将相应的 probit 估计值乘以 0.4。据推测，如果在均值处计算标准逻辑斯蒂函数和 logit 估计值，结果将接近 0.25。

LPM 模型与 logit 和 probit 模型的最大区别在于，LPM 假定 *educ*、*kidslt6* 等自变量的边际效应为常数，而 logit 和 probit 模型则意味着偏效应的大小是递减的。在 LPM 中，估计多一个幼年子女会使劳动市场参与的概率下
539 降约 0.262（无论这个妇女已经有了多少个幼年子女，也不管其他解释变量为多大）。可以将它与 probit 模型估计的边际效应相比较。为简洁起见，找一个 $nwifeinc=20.13$，$educ=12.3$，$exper=10.6$，$age=42.5$（基本上都是样本平均值）和 $kidsge6=1$ 的妇女。估计她从没有幼年子女到有一个幼年子女会使她参加工作的概率下降多少。在其他自变量取上述值的情况下，分别在 $kidslt6=1$ 和 $kidslt6=0$ 处计算标准正态累积密度函数 $\Phi(\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_kx_k)$。我们大致得到 $0.373-0.707=-0.334$，这意味着妇女有一个幼年子女时会降低劳动市场参与的概率约为 0.334。它与换算的 probit 系数 -0.347 相差不大。如果一位妇女的幼年子女数从一个增加到两个，她参与劳动的概率下降得更多，但边际效应并不大：$0.117-0.373=-0.256$。有趣的是，线性概率模型的估计值（被认为是估计了平均值附近的效应）实际上介于这两个估计值之间。

问题 17.2☞

利用 probit 估计值和微积分近似计算，当 *exper* 从 10 提高到 11 时，响应概率的近似变化是多少？

有关线性概率模型中内生解释变量的问题在 logit 和 probit 模型同样出现。虽然我们在这里没有篇幅对它们进行讨论，但利用与两阶段最小二乘相关的方法有可能对内生解释变量进行检验和纠正。埃文斯和斯科韦伯（Evans and Schwab，1995）估计了一个学生是否上大学的 probit 模型，其中关键的解释变量是一个表明该学生是否就读于教会学校的虚拟变量。埃文斯和斯科韦伯用最大似然法估计了一个模型，以便内生地考虑这个变量。［对这些方法

的解释，可参见 Wooldridge(1999，第 15 章)。]

在 probit 模型的背景下，另两个问题也值得注意。第一个是潜变量模型(17.6)中 e 的非线性。自然地，如果 e 不服从标准正态分布，响应概率就不具有 probit 的形式。虽然有些作者倾向于强调估计 β_j 时的不一致性，但除非我们只关心影响的方向，否则就是把注意点摆错了位置。由于响应概率未知，所以即使得到了 β_j 的一致估计，仍不能估计偏效应的大小。

第二个设定问题(仍用潜变量模型中的术语定义)是 e 的异方差性。若 $\mathrm{Var}(e|\boldsymbol{x})$ 与 $\boldsymbol{x}$ 有关，则响应概率不再具有 $G(\beta_0+\boldsymbol{x\beta})$ 的形式；相反，它取决于方差的形式并要求更一般的估计。由于含有自变量的灵活函数形式的 logit 和 probit 模型能很好地起作用，所以这种模型在实践中并不常用。

对二值响应模型略加修改，就适用于独立混合横截面数据或其他观测上独立但不一定同分布的数据集。为了解释总的时间效应，通常还在模型中包括年度或其他时间段的虚拟变量。恰如线性模型一样，logit 和 probit 模型也可用于在自然实验的背景下评价某些政策的影响。

线性概率模型也可适用于综列数据；典型地，可通过固定效应来估计它
540 (参见第 14 章)。最近，具有观测不到效应的 logit 和 probit 模型越来越流行。这些模型因响应概率的非线性性质而变得复杂，而且难以估计和解释。[参见 Wooldridge(1999，第 15 章)。]

17.2 Tobit 模型

另一类重要的限值因变量，在严格为正值时大致连续，但总体中有一个不可忽略的部分取值为零。个人在某给定月份在酒方面的花费就是一例。在美国 21 岁以上的总体中，这个变量的取值范围很大。但有相当大比例的人，在酒方面的花费为零。下面对 Tobit 模型的讨论省略了对某些细节的验证。[这些验证在 Wooldridge(1999，第 16 章)中给出。]

令 y 表示一个实际上在严格正值域上连续但以正概率取值零的变量。我们完全可以使用 y 的一个线性模型。事实上，线性模型可能是对 $\mathrm{E}(y|x_1,x_2,\cdots,x_k)$ 的良好近似，特别是对均值附近的 x_j 而言。但我们可能会得到负的拟合值，从而导致 y 的预测值为负；这与 LPM 解释二值结果相类似。此外，给定解释变量，对 y 的整个分布有一个估计通常也很有用。

Tobit 模型最容易定义为一个潜变量模型：

$$y^* = \beta_0 + \boldsymbol{x\beta} + u,\ u|\boldsymbol{x} \sim \mathrm{Normal}(0,\sigma^2) \tag{17.15}$$

$$y = \max(0, y^*) \tag{17.16}$$

潜变量 y^* 满足经典线性模型假定；具体而言，它服从具有线性条件均值的正态同方差分布。方程(17.16)意味着，当 $y^* \geq 0$ 时，所观测到的变量 y 等于 y^*，但当 $y^* < 0$ 时，则 $y=0$。由于 y^* 正态分布，所以 y 在严格正值上连

续分布。具体而言，对于正值，给定 x 下 y 的密度与给定 x 下 y^* 的密度一样。而且，u/σ 服从标准正态分布并独立于 x，所以

$$P(y=0|\boldsymbol{x})=P(y^*<0|\boldsymbol{x})=P(u<-\boldsymbol{x\beta})=P(u/\sigma<-\boldsymbol{x\beta}/\sigma)$$
$$=\Phi(-\boldsymbol{x\beta}/\sigma)=1-\Phi(\boldsymbol{x\beta}/\sigma)$$

为了记法上的方便，我们将截距项放到了 x 中。因此，如果 $(\boldsymbol{x}_i, y_i)$ 是得自总体的一次随机抽取，则在给定 $\boldsymbol{x}_i$ 下 y_i 的密度为

$$(2\pi\sigma^2)^{-1/2}\exp[-(y-\boldsymbol{x}_i\boldsymbol{\beta})^2/(2\sigma^2)]=(1/\sigma)\phi[(y-\boldsymbol{x}_i\boldsymbol{\beta})/\sigma],\ y>0 \tag{17.17}$$

$$P(y_i=0|\boldsymbol{x}_i)=1-\Phi(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma) \tag{17.18}$$

式中，ϕ 为标准正态分布函数。

541

我们可以从式(17.17)和式(17.18)得到每个观测 i 的对数似然函数：

$$l_i(\boldsymbol{\beta},\sigma)=1(y_i=0)\log[1-\Phi(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma)]$$
$$+1(y_i>0)\log\{(1/\sigma)\phi[(y_i-\boldsymbol{x}_i\boldsymbol{\beta})/\sigma]\} \tag{17.19}$$

注意这个函数是如何取决于 u 的标准差 σ 和 β_j 的。通过将式(17.19)对 i 求和，就可以得到容量为 n 的一个随机样本的对数似然函数。通过最大化这个对数似然函数，可得到 β 和 σ 的最大似然估计值；这要使用数值方法，尽管在大多数情况下能很容易地由软件例行算出。

问题 17.3 ☞

令 y 表示美国总体中已婚妇女婚外情的次数；我们想用她(特别是她在外面是否有工作)、她丈夫和她家庭的其他特征来解释这个变量。这是一个好的 Tobit 模型吗？

像在 logit 和 probit 模型中一样，每个 Tobit 估计值都有标准误，因此可以用来构造每个 $\hat{\beta}_j$ 的统计量；用于求标准误的矩阵表达式很复杂，这里就不再给出。[例如参见 Wooldridge(1999，第 16 章)。]

用瓦尔德检验或似然比检验很容易对多个排除性约束进行检验。瓦尔德检验的形式与 logit 和 probit 的情形类似；式(17.12)同样给出 LR 检验，其中，我们对受约束模型和不受约束模型当然都是使用 Tobit 对数似然函数。

对Tobit 估计值的解释

利用现代计算机，得到 Tobit 模型的最大似然估计值并不比得到线性模型的 OLS 估计值困难多少，而且二者的结果通常很相似。这就使我们禁不住像解释线性回归的估计值那样解释 Tobit 模型的 $\hat{\beta}_j$。不幸的是，事情并没那么容易。

我们从方程(17.15)看出，β_j 度量了 x_j 对 $E(y^*|\boldsymbol{x})$ 的偏效应，其中 y^* 是潜变量。有时，y^* 也会有有意义的经济含义，但通常都没有。我们想解释的

变量是 y，它是所观测到的结果（比如工作的小时数或慈善捐款数）。例如，作为一个政策问题，我们感兴趣的是，工作的小时数对边际税率变化的敏感度。

我们可从式(17.18)估计出 $P(y=0|\mathbf{x})$，当然我们也能估计出 $P(y>0|\mathbf{x})$。如果我们想估计作为 $\mathbf{x}$ 的函数的 y 的期望值，结果会怎么样呢？在 Tobit 模型中，有两个期望值值得特别注意：基于 $y>0$ 而有时被称为“条件期望”的 $E(y|y>0,\mathbf{x})$，和不基于 $y>0$ 而不幸被称为“无条件期望”的 $E(y|\mathbf{x})$。（其实这两个期望值都以解释变量为条件。）期望值 $E(y|y>0,\mathbf{x})$ 告诉我们，对于给定的 $\mathbf{x}$ 值，y 是在 y 为正值的子总体中的期望值。给定 $E(y|y>0,\mathbf{x})$，我们很容易就得到 $E(y|\mathbf{x})$：

$$\begin{aligned}E(y|\mathbf{x}) &= P(y>0|\mathbf{x})\cdot E(y|y>0,\mathbf{x})\\ &= \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)\cdot E(y|y>0,\mathbf{x})\end{aligned} \tag{17.20}$$

为得到 $E(y|y>0,\mathbf{x})$，我们利用正态分布随机变量的一个结论：若 $z\sim\text{Normal}(0,1)$，则 $E(z|z>c)=\phi(c)/[1-\Phi(c)]$ 对任意常数 c 都成立。但由于 $\phi(-c)=\phi(c)$，$1-\Phi(-c)=\Phi(c)$，且 u/σ 服从独立于 $\mathbf{x}$ 的标准正态分
542 布，所以 $E(y|y>0,\mathbf{x})=\mathbf{x}\boldsymbol{\beta}+E(u|u>-\mathbf{x}\boldsymbol{\beta})=\mathbf{x}\boldsymbol{\beta}+\sigma E[(u/\sigma)|(u/\sigma)>-\mathbf{x}\boldsymbol{\beta}/\sigma]=\mathbf{x}\boldsymbol{\beta}+\sigma\phi(\mathbf{x}\boldsymbol{\beta}/\sigma)/\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$。

我们可以将其总结为

$$E(y|y>0,\mathbf{x})=\mathbf{x}\boldsymbol{\beta}+\sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma) \tag{17.21}$$

其中 $\lambda(c)=\phi(c)/\Phi(c)$ 被称为**逆米尔斯比率**（inverse Mills ratio）；它是标准正态 pdf 和标准正态 cdf 在 c 处的值之比。

方程(17.21)很重要。它表明，y 以 $y>0$ 为条件的期望值等于 $\mathbf{x}\boldsymbol{\beta}$ 与一个严格为正的项之和，这个正项等于 σ 乘以逆米尔斯比率在 $\mathbf{x}\boldsymbol{\beta}/\sigma$ 处的值。这个方程还表明，为什么只对 $y_i>0$ 的观测用 OLS 还不能一致地估计 $\boldsymbol{\beta}$；实质上，逆米尔斯比率是被漏掉的一个变量，并一般与 $\mathbf{x}$ 的元素相关。

合并式(17.20)和式(17.21)就得到

$$\begin{aligned}E(y|\mathbf{x}) &= \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}+\sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\\ &= \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)\mathbf{x}\boldsymbol{\beta}+\sigma\phi(\mathbf{x}\boldsymbol{\beta}/\sigma)\end{aligned} \tag{17.22}$$

其中第二个等式的成立是因为 $\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)=\phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$。这个方程表明，当 y 服从一个 Tobit 模型时，$E(y|\mathbf{x})$ 是 $\mathbf{x}$ 和 $\boldsymbol{\beta}$ 的一个非线性函数，从而使得偏效应难以得到。这正是使用 Tobit 模型的代价之一。

如果 x_j 是一个连续变量，那我们就能通过微分求出偏效应。首先

$$\partial E(y|y>0,\mathbf{x})/\partial x_j=\beta_j+\beta_j\cdot(\mathrm{d}\lambda/\mathrm{d}c)(\mathbf{x}\boldsymbol{\beta}/\sigma)$$

假定 x_j 与其他回归元不存在函数相关。通过将 $\lambda(c)=\phi(c)/\Phi(c)$ 进行微分，并利用 $\mathrm{d}\Phi/\mathrm{d}c=\phi(c)$ 和 $\mathrm{d}\phi/\mathrm{d}c=-c\phi(c)$，可以证明 $\mathrm{d}\lambda/\mathrm{d}c=-\lambda(c)[c+\lambda(c)]$。因此

$$\partial E(y|y>0,\mathbf{x})/\partial x_j=\beta_j\{1-\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma+\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\} \tag{17.23}$$

这表明 x_j 对 $\mathrm{E}(y|y>0,\boldsymbol{x})$ 的偏效应并非仅由 β_j 决定。大括号中是调整因子，它取决于 $\boldsymbol{x}$ 的一个非线性函数 $\boldsymbol{x\beta}/\sigma=(\beta_0+\beta_1x_1+\cdots+\beta_kx_k)/\sigma$。可以证明，这个调整因子严格介于 0～1 之间。实践中，我们可以通过代入 β_j 和 σ 的 MLE 来估计式(17.23)。像 logit 和 probit 模型一样，我们必须代入 x_j 的值，通常是它的均值或其他有意义的值。

诸如弹性之类的所有常见的经济量都能计算出来。比如，以 $y>0$ 为条件，y 对 x_1 的弹性为

$$\frac{\partial \mathrm{E}(y|y>0,\boldsymbol{x})}{\partial x_1}\cdot\frac{x_1}{\mathrm{E}(y|y>0,\boldsymbol{x})} \tag{17.24}$$

(原文漏掉了分子上的 ∂——译者注。)在 x_1 以各种函数形式(包括水平值形式、对数形式和二次函数形式)出现时，都可以计算上式。

543 若 x_1 是一个二值变量，则 $\mathrm{E}(y|y>0,\boldsymbol{x})$ 在 $x_1=1$ 与 $x_1=0$ 时的差就给出了我们所要求的影响。其他离散变量(如子女数量)也可类似处理。

我们可用式(17.22)求 $\mathrm{E}(y|\boldsymbol{x})$ 对连续的 x_j 的偏效应。这个导数解释了，在 $y=0$ 开始的人为什么在 x_j 变化时可能选择 $y>0$：

$$\begin{aligned}\frac{\partial \mathrm{E}(y|\boldsymbol{x})}{\partial x_j}=&\frac{\partial P(y>0|\boldsymbol{x})}{\partial x_j}\cdot\mathrm{E}(y|y>0,\boldsymbol{x})\\&+P(y>0|\boldsymbol{x})\cdot\frac{\partial \mathrm{E}(y|y>0,\boldsymbol{x})}{\partial x_j}\end{aligned} \tag{17.25}$$

因为 $P(y>0|\boldsymbol{x})=\Phi(\boldsymbol{x\beta}/\sigma)$，所以

$$\frac{\partial P(y>0|\boldsymbol{x})}{\partial x_j}=(\beta_j/\sigma)\phi(\boldsymbol{x\beta}/\sigma) \tag{17.26}$$

而且一旦代入 β_j 和 σ 的 MLE 及 x_j 的特定值，我们还可以估计式(17.25)中的每一项。

引人注目的是，当我们将式(17.23)和式(17.26)代入式(17.25)并利用对任何 c 都成立的 $\Phi(c)\lambda(c)=\phi(c)$，得到

$$\frac{\partial \mathrm{E}(y|\boldsymbol{x})}{\partial x_j}=\beta_j\Phi(\boldsymbol{x\beta}/\sigma) \tag{17.27}$$

方程(17.27)使我们能大致比较 OLS 和 Tobit 估计值。OLS 系数是对 $\partial\mathrm{E}(y|\boldsymbol{x})/\partial x_j$ 的直接估计。为了使 Tobit 估计值具有可比性，我们将它们乘以调整因子在 x_j 的均值处的大小 $\Phi(\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}/\hat{\sigma})$。因为这只是标准正态 cdf 上的一个值，所以它总是介于 0～1 之间的。由于 $\Phi(\boldsymbol{x\beta}/\sigma)=\mathrm{P}(y>0|\boldsymbol{x})$，所以方程(17.27)表明，调整因子随着 $\mathrm{P}(y>0|\boldsymbol{x})$ 趋近于 1 而趋近于 1。(在对所有 i 都有 $y_i>0$ 的极端情形中，Tobit 模型与 OLS 得到相同的估计值。)

例 17.2 已婚妇女的年度劳动供给

文件 MROZ.RAW 包括了 753 个已婚妇女在工作小时数方面的数据，其中有 428 个妇女当年在家庭以外工作挣工资，另外 325 个妇女的工作小时数

则为零。而对于那些工作小时数为正的妇女而言，工作的时间范围也相当宽，从 12 小时到 4 950 小时。因此，年工作小时数很适合用 Tobit 模型。我们还用 OLS(使用全部 753 个观测)估计了一个线性模型。结论由表 17.2 给出。

544 **表 17.2　　对年工作小时数的 OLS 和 Tobit 估计**

	因变量：*hours*	
自变量	线性模型 (OLS)	Tobit (MLE)
nwifeinc	3.45	−8.81
	(2.54)	(4.46)
educ	28.76	80.65
	(12.95)	(21.58)
exper	65.67	131.56
	(9.96)	(17.28)
$exper^2$	−0.700	−1.86
	(0.325)	(0.54)
age	−30.51	54.41
	(4.36)	(7.42)
kidslt6	−442.09	−894.02
	(58.85)	(111.88)
kidsge6	−32.78	−16.22
	(23.18)	(38.64)
常数项	1 330.48	965.31
	(270.78)	(446.44)
对数似然值	—	−3 819.09
R-平方	0.266	0.274
$\hat{\sigma}$	750.18	1 122.02

此表具有几个值得注意的特征。首先，Tobit 系数估计值具有与对应 OLS 估计值相同的符号，而且统计显著性也类似。(可能的例外是 *nwifeinc* 和 *kidsge6* 的系数，但其 t 统计量大小相当。)其次，尽管人们禁不住想比较 OLS 和 Tobit 估计值的大小，但并不是很有信息价值。我们必须小心，不要因为 *kidslt6* 的 Tobit 系数大致是 OLS 系数的 2 倍，就认为 Tobit 模型中工作小时数对幼年子女数量的反应要大得多。

我们可以将 Tobit 估计值乘以式(17.23)和式(17.27)中的调整因子(在估计值和均值处的大小)，以得到对条件期望的偏效应。式(17.23)中的因子约为 0.451。例如，在工作小时数为正的条件下，估计多受一年教育(在所有变量都取均值的基础上)会提高预期的工作小时数 0.451(80.65)≈36.4 小时。这比 545 OLS 估计值多少大一些。利用这个近似，多一个幼年子女，会使预期的工作小时数减少 0.451(894.02)≈403.2 小时。当然，这对一个工作小时数不足 403.2 小时的妇女没有什么意义。对 *kidslt6* 的两个不同值分别估计期望值并得到二

者之差，而不是用微积分近似，可能会更好。

式(17.27)中的调整因子(同样在 x_j 的均值处)约为 0.645。因此，每个 x_j 对预期工作小时数影响的大小(即当我们对那些原本就不工作和原本就工作的人同时解释时)，比我们以 $hours>0$ 为条件所得到的影响更大。

我们已报告了线性回归和 Tobit 模型的 R-平方。OLS 的 R-平方和通常一样。而对 Tobit 模型来说，R-平方则是 y_i 和 $\hat{y}_i$ 之间相关系数的平方，其中 $\hat{y}_i=\Phi(\mathbf{x}\hat{\boldsymbol{\beta}}/\hat{\sigma})\mathbf{x}_i\hat{\boldsymbol{\beta}}+\hat{\sigma}\phi(\mathbf{x}\hat{\boldsymbol{\beta}}/\hat{\sigma})$ 是 $\mathrm{E}(y\mid\mathbf{x}=\mathbf{x}_i)$ 的估计值。这个式子是受到如下启发而得出的，即通常 OLS 的 R-平方等于 y_i 与其拟合值之相关系数的平方[参见方程(3.29)]。在像 Tobit 模型这样的非线性模型中，相关系数的 R-平方并不像式(3.28)那样地等于基于对残差平方求和而得到的 R-平方。这是因为，像前面定义的那样，样本中的拟合值和残差 $y_i-\hat{y}_i$ 之间并非不相关。将 R-平方定义为 y_i 和 $\hat{y}_i$ 之相关系数的平方，具有总介于 0～1 之间的优点；而基于对残差求和的 R-平方则不需要这个性质。

我们可以看到，从 R-平方这个度量指标来看，Tobit 条件均值函数对工作小时数数据的拟合多少要好一些，但并不明显。不过，我们应记得，Tobit 估计值的选择并不是为了最大化 R-平方(而是为了最大化对数似然函数)，而 OLS 估计值却是为了得到最高的 R-平方的值。

Tobit 模型中的设定问题

Tobit 模型，特别是式(17.21)和式(17.22)中的期望表达式，实质上依赖于其背后潜变量模型中的正态性和同方差性。当 $\mathrm{E}(y\mid\mathbf{x})=\beta_0+\beta_1x_1+\cdots+\beta_kx_k$ 时，我们从第 5 章知道，y 的条件正态性在无偏性、一致性和大样本推断中不起作用。异方差性不会影响 OLS 的无偏性和一致性，尽管我们为了进行近似推断而必须计算稳健标准误和检验统计量。在一个 Tobit 模型中，如果式(17.15)中的任意一个假定不成立，就很难知道 Tobit 估计值在估计什么东西。不过，在偏离假定不远的情况下，Tobit 模型仍可能会很好地估计对条件均值的偏效应。虽然有可能在式(17.15)中做更一般的假定，但这种模型的估计和解释都要复杂得多。

至少在某些应用中，Tobit 模型有一个潜在重要的局限性，即以 $y>0$ 为条件的期望值与 $y>0$ 的概率有密切联系。这一点从方程(17.23)和(17.26)显而易见。具体而言，x_j 对 $\mathrm{P}(y>0\mid\mathbf{x})$ 的影响，就像对 $\mathrm{E}(y\mid y>0,\mathbf{x})$ 的影响一样，与 β_j 成比例，这两个函数乘以 β_j 都为正，并仅通过 $\mathbf{x}\boldsymbol{\beta}/\sigma$ 而取决于 $\mathbf{x}$。
546 这就排除了某些有意义的可能性。比如，考虑一项人寿保险政策的价值与一个人年龄之间的关系。由于年轻人可能根本就不那么喜欢有人寿保险，所以 $y>0$ 的概率随年龄的增加而提高(至少在到达某一点之前)。以有人寿保险为条件，因为随着人们越来越接近其寿命的终点时人寿保险就越来越不重要，所以这些政策的价值随着年龄的增加而递减。在 Tobit 模型中则不容许

这种可能性。

规范地评价 Tobit 模型是否适当的方法之一，是估计一个 Tobit 模型，其中的二值结果（比方说 w）在 $y>0$ 时取值 1，而在 $y=0$ 时取值 0。于是，从式(17.18)看，w 服从一个 probit 模型，其中 x_j 的系数是 $\gamma_j=\beta_j/\sigma$。这意味着我们可以用 probit 模型对每个 j 估计 β_j 与 σ 的比值。若 Tobit 模型成立，则 probit 估计值应接近于 $\hat{\beta}_j/\hat{\sigma}_j$，其中 $\hat{\beta}_j$ 和 $\hat{\sigma}_j$ 是 Tobit 估计值。虽然它们因抽样误差而不可能完全相同，但我们可以检查某些成问题的符号。比如，若 $\hat{\gamma}_j$ 显著并为负，而 $\hat{\beta}_j$ 为正，则说明 Tobit 模型可能不合适。或者说，若 $\hat{\gamma}_j$ 和 $\hat{\beta}_j$ 的符号相同，但 $|\hat{\beta}_j/\hat{\sigma}_j|$ 远大于或远小于 $|\hat{\gamma}_j|$，这也表明有问题。可是对那些在两个模型中都不显著的解释变量，其符号不同或相差很大都不用过多担心。

在年工作小时数的例子中，$\hat{\sigma}=1\,122.02$。将 *nwifeinc* 的 Tobit 系数除以 $\hat{\sigma}$ 得到 $-8.81/1\,122.02\approx-0.007\,9$；*nwifeinc* 的 probit 系数约为 -0.012，二者有差别，但差别不大。*kidslt6* 的 Tobit 系数除以 $\hat{\sigma}$ 约为 -0.797，与 probit 估计值约为 -0.868 相比，同样不是巨大的差别，但仍表明了，有年幼子女对最初的劳动市场参与决策的影响，比在她已经参与劳动后对其选择劳动小时数的影响更大。（Tobit 模型有效地平均了这两种影响。）我们不知道这些影响在统计上是否不同，但它们的大小相当。

如果我们断定 Tobit 模型不合适，结果会怎么样呢？在 Tobit 模型看起来不适合时，还可以使用一些通常被称为**围栏**(hurdle)模型或**两部分**(two-part)模型的模型。它们具有的共同特点是，$\mathrm{P}(y>0|\boldsymbol{x})$ 和 $\mathrm{E}(y|\boldsymbol{x},y>0)$ 取决于不同的参数，所以 x_j 对这两个函数可能具有极为不同的影响。[对这些模型的介绍，可参见 Wooldridge (1999，第 16 章)。]

17.3 泊松回归模型

另一类非负因变量是**计数变量**(count variable)，它可以取非负整数值 $\{0,1,2,\cdots\}$。我们特别感兴趣的是，y 只取包括 0 在内有限几个值的情况。这样的例子有妇女曾生育的子女数、某人在某年被拘捕的次数或一个企业在某年申请专利的个数等。出于前面针对二值和 Tobit 响应所讨论的同样原因，$\mathrm{E}(y|x_1,x_2,\cdots,x_k)$ 的线性模型恐怕不能对所有解释变量的值提供最好的拟合。（不过，像我们在例 3.5 中所做的一样，从一个线性模型开始总是有价值的。）

像 Tobit 结果一样，因为计数变量取值为零，所以我们不能对它取对数，一个有价值的方法是将期望值模型化为一个指数函数：

$$\mathrm{E}(y|x_1,x_2,\cdots,x_k)=\exp(\beta_0+\beta_1x_1+\cdots+\beta_kx_k) \tag{17.28}$$

547 由于 $\exp(\cdot)$ 总为正，所以式(17.28)确保了 y 的预测值也总为正。

尽管式(17.28)比一个线性模型更复杂，但我们基本上已经知道如何去

解释其系数。将方程(17.28)取对数表明

$$\log[\mathrm{E}(y|x_1, x_2, \cdots, x_k)] = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{17.29}$$

所以期望值的对数是线性的。因此，利用我们在前面章节中经常使用的对数函数的近似特征，有

$$\%\Delta\mathrm{E}(y|\boldsymbol{x}) \approx (100\beta_j)\Delta x_j$$

换言之，给定 x_j 提高一个单位，$100\beta_j$ 大致表示了 $\mathrm{E}(y|\boldsymbol{x})$ 变化的百分数。有时需要一个更精确的估计值，我们通过检查期望值的离散变化很容易得到一个这样的估计值。保持除 x_k 外所有的解释变量不变，并令 x_k^0 为其初始值，x_k^1 为其后来的值。于是，期望值的比例变化为

$$\frac{\exp(\beta_0 + \boldsymbol{x}_{k-1}\boldsymbol{\beta}_{k-1} + \beta_k x_k^1)}{\exp(\beta_0 + \boldsymbol{x}_{k-1}\boldsymbol{\beta}_{k-1} + \beta_k x_k^0)} - 1 = \exp(\beta_k \Delta x_k) - 1$$

式中，$\boldsymbol{x}_{k-1}\boldsymbol{\beta}_{k-1}$ 为 $\beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1}$ 的简记，$\Delta x_k = x_k^1 - x_k^0$。当 $\Delta x_k = 1$(比如是一个虚拟变量，我们把它从 0 变化到 1)时，这个变化量就是 $\exp(\beta_k) - 1$。给定 $\hat{\beta}_k$，我们得到 $\exp(\hat{\beta}_k) - 1$，将它乘以 100 后，就把比例变化转化成了变化的百分数。

根据类似线性模型的逻辑，若 β_j 乘以 $\log(x_j)$，则 β_j 就是一个弹性。在实践中，我们起码可以像将 $\log(y)$ 作为因变量的线性模型那样解释方程(17.28)中的系数。其中也有一些我们在此不必深究的微妙差别。

因为方程(17.28)是其参数的一个非线性函数[记住 exp(·)是一个非线性函数]，所以我们不能使用线性回归模型。我们可以使用像 OLS 那样最小化残差平方和的**非线性最小二乘法**。然而，我们发现所有标准计数数据的分布都表现出异方差性，而非线性最小二乘又没有利用到这一点[参见 Wooldridge(1999，第 12 章)]。因而，我们要依赖于最大似然估计和**准最大似然估计**这个有关的重要方法。

在第 4 章，我们把正态性作为线性回归的标准分布假定而引进。正态性假定对取值范围很大的(大致)连续因变量而言是合理的。计数变量不可能具有正态分布(因为正态分布是能取所有值的连续变量)，而且如果它只取很少的几个值，那么这个分布与正态分布就相差很远。对计数数据来说，令人满意的分布则是**泊松分布**(Poisson distribution)。

由于我们感兴趣的是解释变量对 y 的影响，所以我们必须看一下以 $\boldsymbol{x}$ 为条件的泊松分布。因为泊松分布完全由其均值决定，所以我们只须确定 $\mathrm{E}(y|\boldsymbol{x})$。我们假定它具有与式(17.28)同样的形式，简记为 $\exp(\boldsymbol{x\beta})$。于是，以 $\boldsymbol{x}$
548 为条件，y 等于 h 的概率为

$$\mathrm{P}(y = h|\boldsymbol{x}) = \exp[-\exp(\boldsymbol{x\beta})][\exp(\boldsymbol{x\beta})]^h/h!,\ h = 0, 1, \cdots$$

式中，$h!$ 为阶乘(参见附录 B)。这个作为**泊松回归模型**(Poisson regression model)基础的分布，使我们能求出对应于解释变量所有值的条件概率。比如，$\mathrm{P}(y = 0|\boldsymbol{x}) = \exp[-\exp(\boldsymbol{x\beta})]$。一旦我们有了 β_j 的估计值，就可以代入其中

求出对于 x 各个值的概率。

给定一个样本 $\{(\boldsymbol{x}_i, y_i): i=1,2,\cdots,n\}$，我们可以构造对数似然函数：

$$\mathcal{L}(\boldsymbol{\beta})=\sum_{i=1}^{n}\ell_i(\boldsymbol{\beta})=\sum_{i=1}^{n}\{y_i\boldsymbol{x}_i\beta-\exp(\boldsymbol{x}_i\boldsymbol{\beta})\} \tag{17.30}$$

其中我们去掉了 $-\log(y_i!)$ 一项，因为它与 $\boldsymbol{\beta}$ 无关。最大化这个对数似然函数很简单，尽管找不到泊松 MLE 的最终形式。

在最大化对数似然函数之后，就容易得到泊松估计值 $\hat{\beta}_j$ 的标准误；见本章附录中的公式。任何一个软件包都会把它们与 $\hat{\beta}_j$ 一起报告出来。

尽管泊松 MLE 分析是对计数数据分析的自然起步，但它常常受到诸多限制。泊松分布的所有概率以及更高阶矩完全由其均值决定。特别是，方差等于均值：

$$\mathrm{Var}(y|\boldsymbol{x})=\mathrm{E}(y|\boldsymbol{x}) \tag{7.31}$$

这一点很有局限性，而且事实表明，在多数应用中都与式(7.31)不符。幸运的是，泊松分布具有一个很好的稳健性质：不管泊松分布成立与否，我们仍能得到 β_j 的一致和渐近正态的估计量。（这与 OLS 估计量相似，无论正态性假定成立与否，它都是一致和渐近正态的；只是 OLS 在正态性假定下还是 MLE。）[至于细节，可参见 Wooldridge(1999,第 19 章)。]

当我们在没有假定泊松分布完全正确的情况下使用泊松 MLE 时，我们称之为**准最大似然估计**(quasi-maximum likelihood estimation, QMLE)。由于泊松 QMLE 在许多计量经济软件包中都编有程序，所以使用起来很方便。不过，除非泊松方差假定式(7.31)成立，否则还要对标准误进行调整。

当我们假定方差与均值成比例时，可对标准误进行简单调整：

$$\mathrm{Var}(y|\boldsymbol{x})=\sigma^2\mathrm{E}(y|\boldsymbol{x}) \tag{7.32}$$

式中，$\sigma^2>0$ 为未知参数。当 $\sigma^2=1$ 时，我们就得到泊松方差假定。当 $\sigma^2>1$ 时，方差对所有的 $\boldsymbol{x}$ 都大于均值；由于此时方差比泊松情形的方差大(在计数回归的大多数应用中都可观察到这一点)，所以被称为**过度分散**(overdispersion)。被称为**分散不足**(underdispersion)的情形 $\sigma^2<1$ 则不那么常见，但在式(7.32)中也是容许的。549

在式(7.32)下，调整通常的泊松 MLE 标准误很容易。令 $\hat{\beta}_j$ 表示泊松 QMLE，并定义残差为 $\hat{u}_i=y_i-\hat{y}_i$，其中 $\hat{y}_i=\exp(\hat{\beta}_0+\hat{\beta}_1x_{i1}+\cdots+\hat{\beta}_kx_{ik})$ 为拟合值。和往常一样，第 i 次观测的残差就是 y_i 与其拟合值之差。σ^2 的一个一致估计量是 $(n-k-1)^{-1}\sum_{i=1}^{n}\hat{u}_i^2/\hat{y}_i$，其中除以 $\hat{y}_i$ 是进行适当的异方差调整，而 $n-k-1$ 是给定 n 个观测和 $k+1$ 个估计值 $\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_k$ 时的自由度。令 $\hat{\sigma}$ 为 $\hat{\sigma}^2$ 的正平方根，我们将通常的泊松标准误乘以 $\hat{\sigma}$。若 $\hat{\sigma}$ 明显大于 1，则修正的标准误可以比象征性的(一般不正确)泊松 MLE 标准误大很多。

即便式(17.32)也不完全是一般性的。恰如在线性模型中一样，我们可以得到完全不限制方差的泊松 QMLE 的标准误。[进一步的解释，可参见 Wooldridge(1999)。]

问题 17.4 ☞ 假设我们得到 $\hat{\sigma}^2=2$。调整标准误与通常的泊松 MLE 标准误相比如何？准 *LR* 统计量与通常的 *LR* 统计量相比如何？

在泊松分布下，我们可用似然比统计量来检验排除性约束，这个统计量总具有式(17.12)中的形式。如果我们有 q 个排除性约束，那么这个统计量在虚拟假设下就近似服从 χ_q^2 分布。在不那么有限制性的假定(17.32)下，可进行一种简单调整[然后我们就称这个统计量为**准似然比统计量**(quasi-likelihood ratio statistic)]：将式(17.12)除以 $\hat{\sigma}^2$，其中 $\hat{\sigma}^2$ 从无约束模型中得到。

例 17.3 拘捕次数的泊松回归

我们现在将有许多用处的泊松回归模型用于例 9.1 中的拘捕数据。因变量 *narr*86 是一个人在 1986 年被拘捕的次数。这个变量对样本 2 725 个人中的 1 970 个人都是零，而且只有 8 个 *narr*86 的值大于 5。因此，泊松回归模型比线性回归模型更适合。表 17.3 还给出了线性回归模型 OLS 估计值的结论。

550

表 17.3 影响年轻人被拘捕次数的决定因素

因变量：*narr*86		
自变量	线性模型 (OLS)	指数 (泊松 QMLE)
pcnv	-1.32	-0.402
	(0.040)	(0.085)
avgsen	-0.011	-0.024
	(0.012)	(0.020)
tottime	0.012	0.024
	(0.009)	(0.015)
*ptime*86	-0.041	-0.099
	(0.009)	(0.021)
*qemp*86	-0.051	0.038
	(0.014)	(0.029)
*inc*86	-0.001 5	-0.008 1
	(0.000 3)	(0.001 0)
black	0.327	0.661
	(0.045)	(0.074)
hispan	0.194	0.500
	(0.040)	(0.074)
*born*60	-0.022	-0.051
	(0.033)	(0.064)
常数项	0.577	-0.600
	(0.038)	(0.067)
对数似然值	—	2 248.76
R-平方	0.073	0.077
$\hat{\sigma}$	0.829	1.232

OLS 的标准误是通常的标准误；我们当然可以使之对异方差性稳健。泊松回归的标准误是通常最大似然标准误。因为 $\hat{\sigma}=1.232$，所以泊松回归的标准误应该乘以这个因子(因此每个修正后的标准误约高出 23%)。比如，*tottime* 的一个更可靠的标准误是 $1.23(0.015)\approx 0.018\,5$，相应的 t 统计量约为 1.3。虽然对标准误的调整使所有变量的显著性下降，但它们中有几个仍是统计显著的。

OLS 和泊松系数并不直接可比，而且它们具有极为不同的含义。例如，*pcnv* 的系数意味着，若 $\Delta pcnv=0.10$，则期望被拘捕的次数下降 0.013 次(*pcnv* 是先前被拘捕后被定罪的比例)。泊松系数则意味着，$\Delta pcnv=0.10$ 降低预期拘捕约 4%[$0.402(0.10)=0.040\,2$，乘以 100 后就得到百分数影响]。作为一个政策问题，这表明，如果我们提高被定罪的概率 0.10，就能使总拘捕次数下降 4%。

black 的泊松系数意味着，在其他因素不变的情况下，一个黑人被拘捕的
531 次数比一个白人预期高出约 66%。这个系数是高度统计显著的，*hispan* 的系数也一样。

就像在例 17.2 中对 Tobit 模型的应用一样，我们这里也报告了泊松回归的一个 R-平方。它是 y_i 与 $\hat{y}_i=\exp(\hat{\beta}_0+\hat{\beta}_1x_{i1}+\cdots+\hat{\beta}_kx_{ik})$ 之间相关系数的平方。使用这个拟合优度度量的动机与 Tobit 模型一样。我们看到，用泊松 QMLE 估计的指数回归模型拟合得略微好些。记住，OLS 估计值的选择是为了最大化 R-平方，但泊松估计值则不是。(泊松估计值还使对数似然函数最大化。)

在实践中，还有人提出并使用了其他的计数数据回归模型，这些以各种方式推广了泊松分布。如果我们感兴趣的是 x_j 对平均响应的影响，那就没有什么理由不用泊松回归：它简单，通常能给出好的结论并具有前面讨论的稳健性质。实际上，假定式(17.28)成立，我们可以将泊松回归用于 Tobit 一类结果的 y。这可能会给出对平均影响的很好的估计值。当我们对估计诸如 $\mathrm{P}(y>1\mid x)$ 之类的概率感兴趣时，推广泊松回归就更有用。[比如，参见 Cameron 和 Trivedi(1998)。]

17.4 截取和断尾回归模型

一个与 Tobit 模型有类似统计结构的模型被称为**截取回归模型**(censored regression model)。尽管在计量经济学中“Tobit”与“截取回归”经常交替使用，但实际上它们之间有一个重要的区别。Tobit 模型适用于在正值上大致连续分布但以正概率取值零的**结果变量**(outcome variables)。我们已在例 17.2 中已婚妇女的劳动参与情形中看到了这方面的例子，并讨论了诸如慈善捐款数量之类的其他例子。与 Tobit 模型不同，截取回归模型是**因数据截取**(data censoring)而产生。具体而言，其背后的因变量是大致连续的(而且我们将假定它遵循以解释变量为条件的正态分布)，但由于数据搜集方式或制度约束方面的问题，将因变量低于或高于某特定值的部分截取掉。在某种意义上，用截取回归解决的问题是一个数据缺失的问题，只是我们拥有关于缺失数据性质的有用信息。

当我们在抽样方案中以 y 为依据排除了总体的一个子集时，就出现了**断尾回归模型**(truncated regression model)。换句话说，虽然我们没有潜在总体的一个随机样本，但我们知道有哪些单位被包括进样本中来的规则。这个规则由 y 是否高于或低于某个特定的临界值来决定。我们以后还会更详尽地解释截取回归和断尾回归模型的区别。

截取回归模型

552 尽管截取回归模型无须借助于分布的假定而定义，但我们在本小节还是研究**截取正态回归模型**(censored normal regression model)。我们想要解释的变量 y 服从经典线性模型。为强调概念，我们在从总体的一个随机抽取上加下标 i：

$$y_i = \beta_0 + \boldsymbol{x}_i\boldsymbol{\beta} + u_i,\quad u_i \mid \boldsymbol{x}_i,\ c_i \sim \text{Normal}(0, \sigma^2) \tag{17.33}$$

$$w_i = \min(y_i,\ c_i) \tag{17.34}$$

我们不观测 y_i，只有在它小于截取值 c_i 时才能把它观测到。注意式(17.33)包括了 u_i 独立于 c_i 的假定(至少在以 $\boldsymbol{x}_i$ 为条件时)。(为简明起见，我们明确考虑从上截取或**右端截取**；而从下截取或**左端截取**的问题可类似处理。)

问题 17.5 ☞

令 mvp_i 表示工人 i 的边际价值产品；它是一个企业产品的价格与这个工人的边际产品之积。假定 mvp_i 是诸如受教育程度、工作经历等外生变量及不可观测误差的一个线性函数。在完全竞争和没有制度约束的情况下，每个工人所得到的工资都是他或她的边际价值产品。令 $minwage_i$ 表示工人 i 的最低工资，并因其所处的州不同而不同。我们观测 mvp_i 和 $minwage_i$ 的较大者作为 $wage_i$。写出观测工资的适当模型。

右端数据截取的一个例子是**顶端编码**(top coding)。当一个变量超过了顶端编码时，我们只知道它达到了某个临界值。对于高于这个临界值的回答，我们只知道这个变量至少和临界值一样大。比如，在某些调查中，家庭财富有顶端编码。假设受访者被问及其财富状况，通常允许他们回答“高于 50 万美元”。于是，我们可以观察到那些财富不足 50 万美元的受访者的实际财富，但不能观察到那些财富高于 50 万美元的受访者的实际财富。在这种情况下，截取临界值 c_i 对所有的 i 都是一样的。在多数情况下，截取临界值随着个人和家庭的特征而变化。

如果我们观察到($\boldsymbol{x}$, y)的一个随机样本，那我们就很简单地用 OLS 估

计 $\boldsymbol{\beta}$，而且统计推断也是标准程序。（为简单起见，我们又把截距项放到 $\boldsymbol{x}$ 中。）截取将带来问题。利用与 Tobit 模型相似的论证，只用截取观测（即 $y_i < c_i$）的 OLS 回归导致估计量的不一致性。除非不存在截取问题，否则 w_i 对 x_i 的、利用所有观测的 OLS 回归就不会一致地估计 β_j。这与 Tobit 模型类似，但问题相当不同。在 Tobit 模型中，我们是为结果常常为零的经济行为建立模型；并且 Tobit 模型假定反映了这一点。而在截取回归中，由于某种原因，数据被截取了，所以我们遇到一个数据选择问题。

在假定(17.33)和(17.34)之下，给定$(\boldsymbol{x}_i, w_i)$的一个随机样本，我们可用最大似然法估计 $\boldsymbol{\beta}$(和 σ^2)。为此，给定$(\boldsymbol{x}_i, c_i)$，我们需要 w_i 的密度函数。对于未截取的观测，$w_i = y_i$，而且 w 的密度与 y_i 的密度 $\mathrm{Normal}(\boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)$ 相同。而对于被截取的观测，我们需要 w_i 在给定 $\boldsymbol{x}_i$ 下等于截取值 c_i 的概率：

$$\begin{aligned}\mathrm{P}(w_i = c_i \mid \boldsymbol{x}_i) &= \mathrm{P}(y_i \geqslant c_i \mid \boldsymbol{x}_i) = \mathrm{P}(u_i \geqslant c_i - \boldsymbol{x}_i\boldsymbol{\beta}) \\ &= 1 - \Phi[(c_i - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma]\end{aligned}$$

553 我们可以将这两部分合并，以得到 w_i 在给定 $\boldsymbol{x}_i$ 和 c_i 下的密度函数：

$$f(w \mid \boldsymbol{x}_i, c_i) = 1 - \Phi[(c_i - \boldsymbol{x}_i\beta)/\sigma], \quad w = c_i \tag{17.35}$$

$$= (1/\sigma)\phi[(w - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma], \quad w < c_i \tag{17.36}$$

通过将每个 i 的密度函数取自然对数，就得到观测 i 的对数似然函数。将这些对数似然函数对 i 求和，我们就能将这个和对 β_j 和 σ 最大化，而得到 MLE。

须知道，在随机抽样的情况下，我们可以像在线性回归模型中那样解释 β_j。这与应用 Tobit 模型（其中我们所关心的期望值是 β_j 的非线性函数）时很不相同。

截取回归模型的一个重要应用是**持续期间分析**(duration analysis)。持续期间是一个度量某事件发生之前持续时间的变量。比如，我们可能想解释一个从监狱释放的重罪犯下次被捕前持续的天数。对于某些重罪犯，这种情况可能再不会发生，或者要经过很长很长的时间之后，以至我们在分析数据时不得不对持续期间进行截取。

在截取正态回归的持续期间应用中，和在顶端编码应用中一样，我们常常使用自然对数作为因变量，这意味着我们对式(17.34)中的截取临界值也取对数。如我们在整章中所见，利用因变量的对数变换，可能会引起将参数解释为百分比变化的问题。此外，由于取对数的多数都是正变量，所以持续期间变量的对数明显比持续期间变量本身更接近正态分布。

例 17.4 重罪犯的持续期间分析

文件 RECID.RAW 包含的数据是，北卡罗来纳监狱中的犯人在释放后到再次被捕所持续的月数，称之为 *durat*。有些犯人在狱中参加了工作培训。我们还控制一系列人口变量及对监狱和犯罪历史的度量。

在 1 445 个犯人中，有 893 人在追踪的持续期间内未被捕；因此，这些观测要被截取掉。截取时间因人而异，从 70 个月到 81 个月不等。

表 17.4 给出了对 log(*durat*)进行截取正态回归的结果。每个系数乘以 100，都表示在其他条件不变的情况下，对应解释变量每提高一个单位，估计预期持续期间变化的百分数。

表 17.4 中有几个系数很有意思。变量 *priors*(以前被定罪的次数)和 *tserved*(在监狱里度过的总月数)对直至下次被拘捕之前的持续时间的影响都是负的。这表明这些变量对犯罪活动有坏的影响而不是起到阻碍作用。比如，此前多被定一次罪，使得到下次被拘捕的时间几乎少 14%。多服役一年使这个持续期间约减少 $100\cdot 12\cdot(0.019)=22.8\%$。多少有些令人吃惊的发现是，一个因重罪服刑的人与一个不是因重罪而服刑的人相比，估计其预期持续期间要长差不多 56%$[\exp(0.444)-1\approx 0.56]$。

554 **表 17.4　　累犯的截取回归估计**

	因变量：log(*durat*)
自变量	
workprg	−0.063
	(0.120)
priors	−0.137
	(0.021)
tserved	−0.019
	(0.003)
felon	0.444
	(0.145)
alcohol	−0.635
	(0.144)
drugs	−0.298
	(0.133)
black	−0.543
	(0.117)
married	0.341
	(0.140)
educ	0.023
	(0.025)
age	0.003 9
	(0.000 6)
常数项	4.099
	(0.348)
对数似然值	−1 597.06
$\hat{\sigma}$	1.810

555 那些有吸毒或酗酒史的人，预期到其下次被拘捕的持续期间则明显短些。(变量 *alcohol* 和 *drugs* 都是二值变量。)老人和在监禁持续期间就已经结婚的人，预期到其下次被拘捕的持续期间则明显长些。黑人的累犯持续期间明显短些，约短 42%$[\exp(-0.543)-1\approx-0.42]$。

关键的政策变量 *workprg* 并没有达到理想效果。点估计值是，在其他条件不变的情况下，参加工作培训的人比没有参加工作培训的人相比，估计其再犯的持续期间约短 6.3%。这个系数的 t 统计量很小，所以我们可能会得出结论，认为工作培训没有影响。这可能是一个自选择问题，也可能是指派参加培训的方式所导致的结果。当然，也可能就是因为工作培训是不奏效的。

在这个例子中，关键是对截取作出解释，特别是因为有 62% 的累犯的持续期间都被截取掉了。如果我们直接对整个样本应用 OLS，把截取区间当做没有截取一样，那么系数估计值就明显不一样。实际上，它们都向零缩减。比如，*priors* 的系数就变成 -0.059(se = 0.009)，而 *alcohol* 的系数则变成 -0.262(se = 0.060)。尽管其影响的方向是一样的，但这些变量的重要性大大削弱。截取回归估计值则可靠得多。

还有其他方式度量表 17.4 中每个解释变量对持续期间的影响，而不是只关注对预期持续期间的影响。对现代持续期间分析的讨论超出了本书的范围。[对这方面的介绍，可参见 Wooldridge(1999，第 20 章)。]

如果违背了截取正态回归模型的某个假定(特别是存在异方差性和非正态性)，那么 MLE 一般都是不一致的。这就说明，由于利用非截取样本的 OLS 在不要求正态性和同方差性的情况下能得到一致估计，所以截取的潜在成本也很大。有些方法不要求我们假定一个分布，但它们过于高深。[参见 Wooldridge(1999，第 16 章)。]

断尾回归模型

断尾回归模型与截取回归模型类似，但在一个重要方面有所不同：在断尾回归模型中，我们不能观测到总体中某一段的任何信息。在针对总体的某特定子集进行调查(可能出于成本方面的考虑)，而完全忽略总体的其他部分时，就会典型地出现这种情况。

例如，豪斯曼和怀斯(Hausman and Wise，1977)利用一个负收入税实验的数据来研究收入的各种决定因素。一个家庭的收入必须低于 1967 年贫困线(贫困线取决于家庭规模)的 1.5 倍才会包括在研究中。

断尾正态回归模型从一个满足经典线性模型假定的潜在总体模型开始：

$$y=\beta_0+\boldsymbol{x}\boldsymbol{\beta}+u,\quad u\mid\boldsymbol{x}\sim\text{Normal}(0,\sigma^2) \tag{17.37}$$

556 记住，这是一个很强的假定集，因为 u 不仅不能与 x 相关，而且还要正态分

布。我们考虑这个模型是因为要放松它有困难。

在式(17.37)下我们知道，给定总体的一个随机样本，OLS是最有效的估计程序。问题出在我们不能观测到总体的一个随机样本：假定MLR.2不成立。具体而言，一个随机抽取(x_i, y_i)只有在$y_i \leqslant c_i$时才能被观测到，其中c_i是可以取决于外生变量(特别是x_i)的断尾临界值。(在豪斯曼和怀斯的例子中，c_i取决于家庭规模。)这意味着，如果$\{(x_i, y_i): i=1, \cdots, n\}$是我们所观测到的样本，那$y_i$就一定小于或等于$c_i$。这不同于$y_i$可以大于$c_i$的截取回归模型；如果$y_i > c_i$，我们不观测$y_i$就是了。在截取回归模型中，我们对所有随机抽取做的观测都观测了$\boldsymbol{x}_i$；而在断尾模型中，我们只在$y_i \leqslant c_i$时才观测到$\boldsymbol{x}_i$。

为了(连同σ)估计β_j(连同σ)，我们需要y_i在给定$y_i \leqslant c_i$和$\boldsymbol{x}_i$下的分布。这个分布可写成

$$g(y \mid \boldsymbol{x}_i, c_i) = \frac{f(y \mid \boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)}{F(c_i \mid \boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)}, y \leqslant c_i \tag{17.38}$$

式中，$f(y \mid \boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)$为均值为$\beta_0 + \boldsymbol{x}_i\boldsymbol{\beta}$和方差为$\sigma^2$的正态分布；$F(c_i \mid \boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)$则是具有同样均值和方差的在$c_i$处的取值的正态cdf。这个以$y_i \leqslant c_i$为条件的密度表达式符合直觉：它是给定$\boldsymbol{x}$下$y$的总体密度除以$y_i$小于或等于$c_i$的概率$\mathrm{P}(y_i \leqslant c_i \mid \boldsymbol{x}_i)$。事实上，我们通过除以在$f(\cdot \mid \boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)$之下$c_i$以左的面积而将密度重新正规化(指规范为pdf——译者注)。

如果我们将式(17.38)取对数，然后对i求和并对β_j和σ^2最大化这个和，我们就得到最大似然估计量。由此得到一致的渐近正态的估计量，包括标准误和对数似然统计量在内的推断也都是标准的。

如果我们把例17.4中被截取的观测的所有数据都去掉，我们就可以把它当做一个断尾样本来分析。这将给我们552个来自断尾正态分布的观测，其中的断尾点因i而异。不过，我们无论如何也不能如此分析持续期间数据(或顶端编码数据)，因为它删除了有用的信息。我们知道893个持续期间数据的下界和解释变量，这本身就是有用的信息；截取回归用到这些信息，而断尾回归则没有用到。

在豪斯曼和怀斯(Hausman and Wise，1977)给出的一个更好的例子中，他们强调，将OLS应用于一个从上断尾的样本，一般会导致估计量向零偏误。这一点在直觉上讲得通。假设我们关心的是收入与受教育水平之间的关系。如果我们只观测收入低于某个临界值的人，那我们就剪掉了收入分布的上端，这就倾向于使估计线相对于整个总体中的真实回归线来说而变得平坦。对只有一个解释变量且对每个观测都选择相同断尾点的情况，可参见图17.2。

557 像在截取回归中一样，如果违背了式(17.37)中的同方差正态假定，那么断尾正态的MLE就是有偏误和不一致的。也有不要求做这些假定的方法，有关的讨论和参考资料，可参见Wooldridge(1999，第17章)。

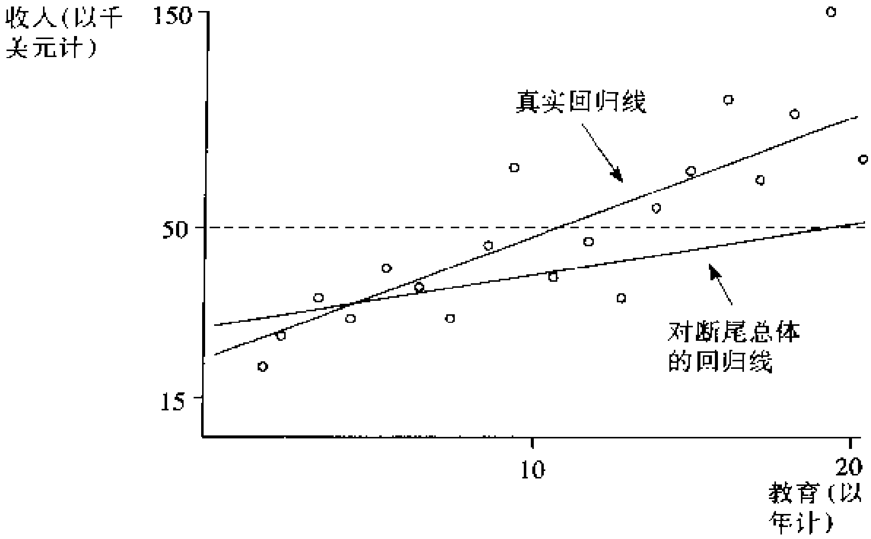


图 17.2 真实或总体回归线与在收入低于 5 万美元处断尾的总体中得到的不正确回归线

17.5 样本选择纠正

断尾回归是所谓**非随机样本选择**(nonrandom sample selection)这个一般问题的特殊情形。但对调查问卷的设计并不是非随机样本选择的惟一原因。回答者常常不能对某些问题作出回答，也会导致因变量和自变量的数据缺失。由于我们在估计中不能利用这些观测，所以我们应该考虑，去掉它们是否使我们的估计量出现偏误。

另一个一般性的例子是通常所谓的偶然断尾(incidental truncation)。这里我们不能观察到 y 是因为另一个变量的结果。有代表性的例子是劳动经济学中估计所谓的**工资报价函数**。人们关注的是，诸如受教育程度等各种因素如何影响劳动市场中工人所挣工资。对于参加劳动市场的人，我们可以观察到其工资报价就是其当前工资。但对那些目前不在劳动市场的人，我们就不能观察到其工资报价。因为有工作可能与观测不到而又影响工资报价的因素系统相关，所以只用有工作的人(到目前为止，我们在所有工资方面的例子中都是这么做)，可能导致工资报价方程中参数估计量产生偏误。 558

非随机样本选择也会出现在综列数据中。在最简单的情形中，我们有两年的数据，但有些人可能因退休等而离开了样本。在人员自然衰减与项目的有效性相关的政策分析中，这就特别成问题。

OLS什么时候对选择的样本是一致的？

我们在9.4节简要地讨论了几种可忽略的样本选择问题，其关键区别在于外生和内生样本选择。在断尾Tobit情形中，我们得到的显然是内生样本选择，因而OLS是有偏误和不一致的。另一方面，如果我们的样本仅由外生解释变量决定，那我们就得到外生样本选择。介于这两个极端之间的情形则不是很明确的，因而我们现在对他们作出小心的定义和假定。总体模型是

$$y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+u,\ \mathrm{E}(u|x_1,\ x_2,\cdots,x_k)=0 \tag{17.39}$$

将一个从中随机抽取的总体模型写成如下形式会有所帮助：

$$y_i=\boldsymbol{x}_i\boldsymbol{\beta}+u_i \tag{17.40}$$

式中，$\boldsymbol{x}_i\boldsymbol{\beta}$ 为 $\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_kx_{ik}$ 的简记。现在，令 n 表示来自总体的一个随机样本的容量。如果我们对所有的 i 都能观测到 y_i 和每个 x_{ij}，那我们就能使用OLS。如果出于某种原因，某个观测 i 的 y_i 或某些自变量不能观测到，但至少对于有些观测，我们能观测到其变量集的全部信息。为每个 i 定义一个选择指标 s_i。若我们观测到 $(y_i,\ \boldsymbol{x}_i)$ 的全部，则 $s_i=1$；否则 $s_i=0$。因此，$s_i=1$ 表示我们在分析中将用到这个观测；$s_i=0$ 则表示不用这个观测。我们感兴趣的是，使用**选择样本**（selected sample）（即使用 $s_i=1$ 的观测）时OLS估计量的统计性质。因此，我们使用观测的个数 n_1 小于 n。

于是很容易得到使OLS成为一致（甚至无偏）估计的条件。事实上，不用估计式(17.40)，我们可以只估计方程

$$s_iy_i=s_i\boldsymbol{x}_i\boldsymbol{\beta}+s_iu_i \tag{17.41}$$

当 $s_i=1$ 时，就得到式(17.40)；当 $s_i=0$ 时，就得到 $0=0+0$，显然没有告诉我们 $\boldsymbol{\beta}$ 的任何信息。将 s_iy_i 对 $s_i\boldsymbol{x}_i(i=1,2,\cdots,\ n)$ 回归，等同于利用 $s_i=1$ 的观测将 y_i 对 $\boldsymbol{x}_i$ 回归。因此，我们可以通过对一个随机样本研究式(17.41)来了解 $\hat{\beta}_j$ 的一致性。

从第5章的分析知道，如果误差项的均值为零并与解释变量无关，则从式(17.41)得到的OLS估计量就是一致的。在总体中，零均值假定是
559 $\mathrm{E}(su)=0$，而零相关假定可表述为

$$\mathrm{E}[(sx_j)(su)]=\mathrm{E}(sx_ju)=0 \tag{17.42}$$

式中，s，x_j 和 u 都是代表总体的随机变量。由于 s 是一个二值变量，所以我们用到 $s^2=s$。如果我们观测到一个随机样本的全部变量，那么条件(17.42)就与我们所需要的条件 $\mathrm{E}(x_ju)=0$ 不同。因此，在总体中，我们需要 u 与 sx_j 无关。

无偏性的关键条件是 $\mathrm{E}(su|sx_1,\cdots,sx_k)=0$。和通常一样，这个假定比一致性所需要的假定更强。

如果 s 仅是解释变量的函数，那么 sx_j 也只是 $x_1,\ x_2,\cdots,x_k$ 的一个函数；

根据式(17.39)中的条件均值假定，sx_j 也与 u 无关。实际上，因为 $\mathrm{E}(u|x_1,x_2,\cdots,x_k)=0$，所以 $\mathrm{E}(su|sx_1,\cdots,sx_k)=s\mathrm{E}(u|sx_1,\cdots,sx_k)=0$。这就是**外生样本选择**(exogenous sample selection)的情形，其中 $s_i=1$ 完全由 $x_{i1},\cdots,x_{ik}$ 决定。举一个例子，如果我们估计一个工资方程，其中的解释变量是受教育程度、工作经历、现职任期、性别、婚姻状况等假定为外生的变量，那么我们就能基于某些或全部解释变量来选择样本。

如果样本选择在 s_i 独立于 $(\mathbf{x}_i, u_i)$ 的意义上是完全随机的，那么因为在式(17.39)下 $\mathrm{E}(x_ju)=0$，所以 $\mathrm{E}(sx_ju)=\mathrm{E}(s)\mathrm{E}(x_ju)=0$。因此，如果我们从一个随机样本开始，并随机地去观测，那么OLS仍是一致的。事实上，给定选择样本中不存在完全共线性，则OLS在这种情形中仍是无偏的。

如果取决于解释变量，附加的随机项又独立于 $\mathbf{x}$ 和 u，OLS也是无偏和一致的。比如，假设 IQ 得分也是工资方程中的一个解释变量，但某些人没有 IQ 数据。假设我们认为，选择可做如下描述：若 $IQ\geqslant v$，则 $s=1$；若 $IQ<v$，则 $s=0$，其中 v 是一个独立于 IQ,u 和其他解释变量的随机变量并观测不到。这意味着我们更可能观测到高的 IQ，但总有某种观测不到任何 IQ 的可能性。以解释变量为条件，s 独立于 u，这意味着 $\mathrm{E}(u|x_1,x_2,\cdots,x_k,s)=\mathrm{E}(u|x_1,x_2,\cdots,x_k)$，且根据对总体模型的假定，最后一个期望值等于零。如果我们增加同方差性假定 $\mathrm{E}(u^2|\mathbf{x},s)=\mathrm{E}(u^2)=\sigma^2$，那么通常的OLS标准误和检验统计量就都成立了。

到目前为止，我们已经看到了几种选择样本的OLS是无偏或至少一致的情形。选择样本的OLS什么时候不一致呢？我们已经看到了一个例子：利用断尾样本的回归。当从上断尾时，若 $y_i\leqslant c_i$，则 $s_i=1$，其中 c_i 是断尾临界值。等价地说，若 $u_i\leqslant c_i-\mathbf{x}_i\boldsymbol{\beta}$，则 $s_i=1$。由于 s_i 直接取决于 u_i，所以 s_i 和 u_i 不可能无关，即使以 $\mathbf{x}_i$ 为条件也不可能。这就是为什么选择样本的OLS不能一致地估计 β_j 的原因。s 与 u 还有方式不那么明显的相关；我们将在下一小节考虑。

关于OLS一致性方面的结论可推广到工具变量估计。如果总体中的IV表示为 z_h，那么2SLS一致性的关键条件是 $\mathrm{E}(sz_hu)=0$，后者在 $\mathrm{E}(u|\mathbf{z},s)=0$
560 时成立。因此，如果选择完全由外生变量 $\mathbf{z}$ 决定，或者 s 取决于独立于 u 和 $\mathbf{z}$ 的其他因素，那么选择样本的2SLS通常都是一致的。我们其实需要假定解释变量和工具变量在总体的被选择部分适当相关。伍德里奇(Wooldridge, 1999, 第17章)的论述包含了对这些假定的精确表述。

还可以证明，当选择完全是外生变量的函数时，非线性模型(如logit和probit模型)的最大似然估计将给出一致和渐近正态的估计量，而通常的标准误和检验统计量也都成立。[同样参见Wooldridge(1999, 第17章)。]

偶然断尾

如我们前面提到的那样，样本选择的常见形式是偶然断尾。我们再次从

式(17.39)中的总体模型开始。不过，假定我们将总能观测到解释变量 x_j。问题是，我们只能观测到总体中 y 的一个子集。我们能否观测到 y 的决定规则并不直接取决于 y 的结果。主要的例子出现于 $y=\log(wage^o)$，其中 $wage^o$ 是工资报价或一个人在劳动市场上能得到的小时工资。如果一个人在被调查时确实在工作，那么由于我们假定了工资是可观测的，所以就能得到其工资报价。但对于那些没有工作的人，我们就不能得到其 $wage^o$。因此，工资报价的断尾就是偶然的，因为它取决于另一个变量(可以说是劳动市场参与变量)。重要的是，我们一般都能观测到个人的所有信息，如受教育水平、先前的工作经历、性别和婚姻状况等。

通常处理偶然断尾问题的方法是，在我们考虑的总体模型中添加一个明确的选择方程：

$$y=\boldsymbol{x\beta}+u,\ \mathrm{E}(u|\boldsymbol{x})=0 \tag{17.43}$$

$$s=1[\boldsymbol{z\gamma}+v\geqslant0] \tag{17.44}$$

式中，若观测到 y，则 $s=1$，否则 $s=0$。我们假定 $\boldsymbol{x}$ 和 $\boldsymbol{z}$ 的各元素都可观测到，记 $\boldsymbol{x\beta}=\beta_0+\beta_1x_1+\cdots+\beta_kx_k$ 和 $\boldsymbol{z\gamma}=\gamma_0+\gamma_1z_1+\cdots+\gamma_mz_m$。

我们主要关心的方程是(17.43)，在给定一个随机样本下可以用 OLS 估计 $\boldsymbol{\beta}$。选择方程(17.44)取决于可观测变量 z_h 和不可观测误差 v。我们将做的一个标准假定是，$\boldsymbol{z}$ 在方程(17.43)中是外生的：

$$\mathrm{E}(u|\boldsymbol{x},\boldsymbol{z})=0$$

实际上，为了让下面提出的方法能很好地起作用，我们将要求 $\boldsymbol{x}$ 严格地是 $\boldsymbol{z}$ 的一个子集；任何一个 x_j 都是 $\boldsymbol{z}$ 的一个元素，而 $\boldsymbol{z}$ 的某些元素则不在 $\boldsymbol{x}$ 中。我们以后会看到这一点为什么很关键。

假定样本选择方程中的误差项 v 与 $\boldsymbol{z}$（因而 $\boldsymbol{x}$)无关，还假定 v 具有标准正态分布。我们很容易看到，u 和 v 之间的相关一般会导致样本选择问题。为了看出其原因，我们假定(u，v)独立于 $\boldsymbol{z}$。于是，以 $\boldsymbol{z}$ 和 v 为条件对方程(17.43)取期望，并利用 $\boldsymbol{x}$ 是 $\boldsymbol{z}$ 的一个子集这一事实，得到

561

$$\mathrm{E}(y|\boldsymbol{z},\ v)=\boldsymbol{x\beta}+\mathrm{E}(u|\boldsymbol{z},\ v)=\boldsymbol{x\beta}+\mathrm{E}(u|v)$$

式中，$\mathrm{E}(u|\boldsymbol{z},\ v)=\mathrm{E}(u|v)$是因为($u$，$v$)独立于 $\boldsymbol{z}$。现在，如果 u 和 v 是联合正态的(且均值为零)，那么就有参数 ρ 使得 $\mathrm{E}(u|v)=\rho v$。因此

$$\mathrm{E}(y|\boldsymbol{z},\ v)=\boldsymbol{x\beta}+\rho v$$

虽然没有观测到 v，但我们可以利用这个方程计算 $\mathrm{E}(y|\boldsymbol{z},\ s)$，然后将它具体化到 $s=1$。我们现在有

$$\mathrm{E}(y|\boldsymbol{z},\ s)=\boldsymbol{x\beta}+\rho\mathrm{E}(v|\boldsymbol{z},\ s)$$

由于根据式(17.44) s 和 v 相关，而且 v 具有标准正态分布，所以当 $s=1$ 时我们可以证明，$\mathrm{E}(v|\boldsymbol{z},\ s)$就是逆米尔斯比 $\lambda(\boldsymbol{z\gamma})$。这就得到一个重要的方程

$$\mathrm{E}(y|\boldsymbol{z},\ s=1)=\boldsymbol{x\beta}+\rho\lambda(\boldsymbol{z\gamma}) \tag{17.45}$$

方程(17.45)表明，给定 $\mathbf{z}$ 及 y 的可观测性，y 的期望值等于 $\mathbf{x}\boldsymbol{\beta}$ 加上取决于逆米尔斯比在 $\mathbf{z}\boldsymbol{\gamma}$ 处值的一项。记住，我们希望估计 $\boldsymbol{\beta}$。这个方程表明，如果将 $\lambda(\mathbf{z}\boldsymbol{\gamma})$ 这一项作为附加回归元包括进来，那我们只用选择的样本就能估计它。

若 $\rho=0$，则 $\lambda(\mathbf{z}\boldsymbol{\gamma})$ 就不会出现，而且 y 对 $\mathbf{x}$ 利用选择样本做 OLS 就能一致地估计 $\boldsymbol{\beta}$。否则，我们实质上省略了一个通常与 $\mathbf{x}$ 相关的变量 $\lambda(\mathbf{z}\boldsymbol{\gamma})$。什么时候 $\rho=0$ 呢？答案是在 u 和 v 不相关时。

由于 $\boldsymbol{\gamma}$ 未知，所以我们不能对每个 i 计算 $\lambda(\mathbf{z}_i\boldsymbol{\gamma})$。但根据我们已做的假定，$s$ 在给定 $\mathbf{z}$ 时服从一个 probit 模型：

$$P(s=1|\mathbf{z})=\Phi(\mathbf{z}\boldsymbol{\gamma}) \tag{17.46}$$

因此，我们可以利用全部样本将 s_i 对 $\mathbf{z}_i$ 做 probit 来估计 $\boldsymbol{\gamma}$，第二步就可以估计 $\boldsymbol{\beta}$ 了。这个程序最近在计量经济学中被称为 Heckit 方法，以纪念赫克曼(Heckman, 1976)的工作，我们对它进行概括如下。

样本选择纠正

(i)利用所有的 n 个观测，估计一个 s_i 对 $\mathbf{z}_i$ 的 probit 模型，并得到估计值 $\hat{\gamma}_h$。对每个 i 计算逆米尔斯比 $\hat{\lambda}_i=\lambda(\mathbf{z}_i\hat{\boldsymbol{\gamma}})$。(实际上，我们只需要对 $s_i=1$ 的 i 做这些。)

(ii)利用选择样本，即 $s_i=1$ 的观测(比方说其中的 n_1 个)，做如下回归

$$y_i \text{ 对 } \mathbf{x}_i, \hat{\lambda}_i \text{ 回归} \tag{17.47}$$

则 $\hat{\beta}_j$ 就是一致的，并近似正态分布。

562 从回归(17.47)中可得到对选择偏误的一个简单检验。即我们可用 $\hat{\lambda}_i$ 通常的 t 统计量作为对 $H_0:\rho=0$ 的一个检验。在 H_0 下，不存在样本选择问题。

当 $\rho\neq0$ 时，回归(17.47)中报告的通常的 OLS 标准误并不完全正确。这是因为，它们没有考虑对 $\boldsymbol{\gamma}$ 的估计使用了回归(17.47)中同样的观测和其他一些观测。有些计量经济软件包计算出了正确的标准误。[不幸的是，它不像对异方差性的调整那么简单。进一步的讨论可参见 Wooldridge(1999，第 6 章)。] 在多数情况下，调整并不会带来重大的不同，但事先很难知道(除非 $\hat{\rho}$ 很小并且不显著)。

我们最近注意到，$\mathbf{x}$ 应该是 $\mathbf{z}$ 的一个严格子集。这有两层含义。首先，式(17.43)中作为解释变量出现的任何一个元素，也应该是选择方程中的一个解释变量。尽管在很少见的情况下，从选择方程中去掉一些元素也讲得通，但在 $\mathbf{z}$ 中包含 $\mathbf{x}$ 中的所有元素，代价也并不大；而如果去掉得不正确的话，则会导致不一致性。

第二个重要含义是，在 $\mathbf{z}$ 中至少有一个元素不在 $\mathbf{x}$ 中。这意味着，我们需要一个影响选择但对 y 没有偏效应的变量。尽管不是必须应用这个程序(实际上，在 $\mathbf{z}=\mathbf{x}$ 时我们可以机械地进行这两步)，但除非我们在式(17.43)中有一个排除性约束，否则结论通常都不是很可信的。其原因是，尽管逆米

尔斯比是 z 的一个非线性函数，但它常常可以用一个线性函数很好地近似。若 $z=x$，则 $\hat{\lambda}_i$ 可能与 x_i 的元素高度相关。如我们所知，这种多重共线性可能导致 $\hat{\beta}_j$ 很高的标准误。直觉上讲，如果我们没有一个影响选择而又不影响 y 的变量，那么，要将式(17.43)中的样本选择与函数形式误设区分开来，若不是不可能，也是极其困难的。

例 17.5　已婚妇女的劳动市场参与

我们对 MROZ.RAW 中已婚妇女的数据进行样本选择纠正。记得样本中的753个妇女中，有428人当年在工作。工资报价方程是以 log(*wage*)为因变量和以 *educ*，*exper* 和 $exper^2$ 作为解释变量的标准方程。为了检验和纠正(因观测不到未工作妇女的工资报价而导致的)样本选择偏误，我们需要估计一个劳动市场参与的 probit 模型。除了受教育水平和工作经历外，我们还包括了表 17.1 中的一些因素：其他收入、年龄、幼年子女个数和年龄较大的子女个数。工资报价方程中对这四个变量的排除只是一个假定：我们假定，给定生产力因素，*nwifeinc*，*age*，*kidslt*6 和 *kidsge*6 对工资报价没有影响。可从表 17.1 中的 probit 结论明显看出，至少 *age* 和 *kidslt*6 对劳动市场参与有很大的影响。

表 17.5 包含了 OLS 和 Heckit 的结论。[Heckit 结论的标准误正是回归(17.47)中通常的标准误。]在工资报价方程中，没有样本选择问题的迹象。$\hat{\lambda}$ 系数的 t 统计量很小(0.239)，所以我们不能拒绝 $H_0: \rho=0$。同样重要的是，563 表 17.5 中估计的斜率系数实际上没有很大的差别。所估计的教育回报只相差1/10个百分点。

表 17.5　　已婚妇女的工资报价方程

因变量：log(*wage*)		
自变量	OLS	Heckit
educ	0.108	0.109
	(0.014)	(0.016)
exper	0.042	0.044
	(0.012)	(0.016)
$exper^2$	−0.000 81	−0.000 86
	(0.000 39)	(0.000 44)
常数项	−0.522	−0.578
	(0.199)	(0.307)
$\hat{\lambda}$	—	0.032
		(0.134)
样本容量	428	428
R-平方	0.157	0.157

进行两步估计方法的另一种办法是完全最大似然估计。由于它要求得到 y 和 s 的联合分布，所以更为复杂。利用前面的程序检验样本选择通常都行得通；如果没有样本选择问题的证据，那就没有必须继续下去。如果我们侦查出样本选择偏误，就可以使用两步估计值或用 MLE 同时估计回归和选择方程。[可参见 Wooldridge(1999，第 17 章)。]

在例 17.5 中，我们不仅仅知道一个女人这一年中是否工作，我们还知道每个女人工作多少小时。事实证明，我们在另一种样本选择程序中可利用这一信息。我们不是使用逆米尔斯比 $\hat{\lambda}_i$，而是使用 Tobit 残差 $\hat{v}_i$，在 $y_i>0$ 时，564 这个残差可计算为 $\hat{v}_i=y_i-\boldsymbol{x}\hat{\boldsymbol{\beta}}$。可以证明，在(17.47)的回归中，以 $\hat{v}_i$ 取代 $\hat{\lambda}_i$ 也能得到 β_j 的一致估计，而且 $\hat{v}_i$ 的 t 统计量也是样本选择偏误的一个有效检验。虽然这种方法具有利用了更多信息的优点，但它没有那么广泛地适用性。[参见 Wooldridge(1999，第 17 章)。]

还有许多关于样本选择的专题。其中值得一提的是在可能的样本选择偏误之外含有内生解释变量的模型。把含有单一个内生解释变量的模型写成

$$y_1=\alpha_1 y_2+\boldsymbol{z}_1\boldsymbol{\beta}_1+u_1 \tag{17.48}$$

式中，y_1 只有在 $s=1$ 时才能观测到，而 y_2 只能与 y_1 一起观测到。举个例子来说，y_1 表示在任者的得票百分比，而 y_2 则表示在任者的竞选支出在竞选总支出中的百分比。对于不参加竞选的在任者来说，就观察不到 y_1 或 y_2。如果我们有一些影响参与竞选的决策而又与竞选支出相关的外生因素，那我们就能用工具变量法估计 α_1 和 $\boldsymbol{\beta}_1$ 的元素。为了令人信服，我们需要两个式(17.48)中没有出现的外生变量。事实上，一个应该影响选举决策，另一个应该与 y_2 相关[用 2SLS 估计方程(17.48)的通常要求]。简单地讲，方法就是用 probit 估计选择方程，其中所有的外生变量都在 probit 方程中出现。然后，我们把逆米尔斯比放到式(17.48)中，再用 2SLS 估计这个方程。由于逆米尔斯比只取决于外生变量，所以它可作为自身的工具变量。我们把所有的外生变量都当做其他的工具变量。和从前一样，我们可以用 $\hat{\lambda}_i$ 的 t 统计量作为对选择偏误的一个检验。[更多的信息，可参见 Wooldridge(1999，第 17 章)。]

▶ 小结

我们在本章探讨了几个在应用研究、特别是在计量经济研究中时常用到的高级方法。logit 和 probit 模型用于二值响应变量。这些模型与线性概率模型相比有一些优点：拟合概率介于 0～1 之间，偏效应递减。logit 和 probit 模型的主要代价是它们更难以解释。

Tobit 模型适用于非负结果，其中大量结果聚集于零，而正值也有宽广的取值范围。许多个人选择变量都具有这个特征，如劳动供给、人寿保险的数量和养老基金投资于股票的数量。与 logit 和 probit 模型一样，给定 $\boldsymbol{x}$，y 的期

望值（以 $y>0$ 为条件或无条件）非线性地取决于 $\boldsymbol{x}$ 和 $\boldsymbol{\beta}$。我们给出这些期望的表达式和每个 x_j 对这些期望的偏效应公式。在估计了 Tobit 模型之后，可以用最大似然法估计这些。

当因变量是一个计数变量（即取非负整数值）时，泊松回归模型则很合适。给定 x_j 下 y 的期望值具有指数形式。这就根据 x_j 是水平值形式或是对数形式而对参数做半弹性或弹性解释。总之，我们可以像在以 log（y）为因变量的线性模型中那样解释参数，也能用 MLE 估计这些参数。不过，由于泊松分布施加了方差等同于均值的条件，所以通常要计算那些容许过度分散或分 565 散不足的标准误和检验统计量。这些都是对 MLE 通常的标准误和统计量的简单调整。

截取和断尾回归模型处理的是特殊形式的数据缺失问题。在截取回归中，因变量在一个临界值之上或之下被截取掉。我们可以利用截取后的结果，因为我们总能观测到解释变量，就像在持续期间分析或对观测做顶端编码的分析中那样。断尾回归模型在总体的一部分被完全排除时出现：我们没有观察到那些抽样方案中没有包括的单元的任何信息。这是样本选择问题之特例。

17.5 节对非随机样本选择进行了系统的讨论。我们证明了，外生样本选择并不影响 OLS 应用于子样本时的一致性，而内生样本选择则有影响。我们还说明了，如何对一般的偶然断尾问题——观测因另一变量（如劳动市场参与）的结果而缺失了 y——检验和纠正其样本选择偏误。在这些情况下，赫克曼的方法相对容易实施。

关键术语

二值响应模型
截取回归模型
角点解
计数变量
数据截取
持续期间分析
外生样本选择
Heckit 模型
偶然断尾
逆米尔斯比
潜变量模型
限值因变量（LDV）
logit 模型
对数似然函数

最大似然估计（MLE）
非随机样本选择
过度分散
正确预测百分数
泊松分布
泊松回归模型
probit 模型
拟 R-平方
拟似然比统计量
拟最大似然估计（QMLE）
选择样本
Tobit 模型
顶端编码
断尾回归模型

习 题

17.1 (i)对于一个二值响应 y，令 $\bar{y}$ 表示样本中为1的比例(等于 y_i 的样本均值)。令 $\hat{q}_0$ 表示结果为 $y=0$的正确预测百分数，而 $\hat{q}_1$ 表示结果为 $y=1$的正确预测百分数。若 $\hat{p}$ 是整体的正确预测百分数，证明 $\hat{p}$ 是 $\hat{q}_0$ 和 $\hat{q}_1$ 的一个加权平均：

566

$$\hat{p} = (1-\bar{y})\hat{q}_0 + \bar{y}\hat{q}_1$$

(ii) 在一个容量为300的样本中，假设 $\bar{y}=0.70$，所以有210个结果为 $y_i=1$，90个结果为 $y_i=0$。假设 $y=0$ 的正确预测百分数为80，而 $y=1$ 的正确预测百分数为40，求总体正确预测百分数。

17.2 令 *grad* 表示一个学生在一所大学是否5年内毕业的虚拟变量。令 *hsGPA* 和 *SAT* 表示高中时的平均成绩和 *SAT* 分数。令 *study* 表示每周在有组织的教学大楼里学习的小时数。假设利用420个学生运动员的数据得到如下 logit 模型：

$$\begin{aligned}\hat{P}(grad &= 1 \mid hsGPA,\ SAT,\ study) \\ &= \Lambda(-1.17 + 0.24\ hsGPA + 0.000\,58\ SAT + 0.073\ study)\end{aligned}$$

式中，$\Lambda(z)=\exp(z)/[1+\exp(z)]$为 logit 方程。保持 *hsGPA* 固定在3.0和 *SAT* 固定在1 200的水平上，计算每周花10小时的同学与每周花5小时的同学在毕业概率上的估计差异。

17.3 (要求一些微积分知识) (i)在 Tobit 模型中假设 $x_1=\log(z_1)$，而且这是 $\boldsymbol{x}$ 中惟一出现 z_1 的地方。证明

$$\frac{\partial \mathrm{E}(y|y>0,\boldsymbol{x})}{\partial z_1} = (\beta_1/z_1)\{1-\lambda(\boldsymbol{x\beta}/\sigma)[\boldsymbol{x\beta}/\sigma+\lambda(\boldsymbol{x\beta}/\sigma)]\} \qquad (17.49)$$

式中，β_1 为 $\log(z_1)$的系数。

(ii)若 $x_1=z_1$ 和 $x_2=z_1^2$，证明

$$\frac{\partial \mathrm{E}(y|y>0,\boldsymbol{x})}{\partial z_1} = (\beta_1+2\beta_2 z_1)\{1-\lambda(\boldsymbol{x\beta}/\sigma)[\boldsymbol{x\beta}/\sigma+\lambda(\boldsymbol{x\beta}/\sigma)]\}$$

式中，β_1 和 β_2 分别为 z_1 和 z_1^2 的系数。

17.4 令 mvp_i 表示工人 i 的边际价值产品，即企业产品的价格与该工人边际产品的乘积。假定

$$\begin{aligned}&\log(mvp_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i \\ &wage_i = \max(mvp_i,\ minwage_i)\end{aligned}$$

式中，解释变量包括受教育水平、工作经历等，而 $minwage_i$ 是第 i 个人适当

的最低工资。用 $\log(mvp_t)$ 和 $\log(minwage_t)$ 表示 $\log(wage_t)$。

17.5 （要求一些微积分知识）令 *patents* 表示一个企业在给定年份申请专利的个数。假定给定 *sales* 和 *RD* 下 *patents* 的条件期望为

$$\mathrm{E}(patents \mid sales,\ RD) = \exp[\beta_0 + \beta_1 \log(sales) + \beta_2 RD + \beta_3 RD^2]$$

式中，*sales* 为企业的年销售量；*RD* 为在过去 10 年间在研发方面的总支出。

567 (i) 你将如何估计 β_j？通过讨论 *patents* 的性质说明你的回答是正确的。

(ii) 你如何解释 β_1？

(iii) 求出 *RD* 对 $\mathrm{E}(patents \mid sales,\ RD)$ 的偏效应。

17.6 对美国所有家庭构成的总体考虑一个家庭储蓄方程：

$$sav = \beta_0 + \beta_1 inc + \beta_2 hhsize + \beta_3 educ + \beta_4 age + u$$

式中，*inc* 为家庭收入；*hhsize* 为家庭规模；*educ* 为户主受教育年数；*age* 为户主的年龄。假定 $\mathrm{E}(u \mid inc,\ hhsize,\ educ,\ age) = 0$。

(i) 假设样本只包括户主年龄在 25 岁以上的家庭。如果我们对这样一个样本使用 OLS，能得到 β_j 的无偏估计量吗？请解释。

(ii) 现在假设样本只包括无子女的已婚夫妇。我们能估计储蓄方程中的所有参数吗？能估计哪些参数？

(iii) 假设从样本中排除掉储蓄超过每年 25 000 美元以上的家庭。OLS 能得到 β_j 的一致估计量吗？

17.7 假设你被一所大学雇用，任务是研究决定一个申请入读该校的学生实际上是否会入读的因素。他们给了你去年申请学生的一个很大的随机样本。你还有每个学生是否入读、高中表现、家庭收入、得到的助学援助、种族和地理变量。有人对你说："对这些数据的任何分析都会得到有偏误的结论，因为它不是所有大学（而只是这个大学）申请者的一个随机样本。"你怎么看待这种批评？

计算机习题

17.8 本题利用 PNTSPRD.RAW 中的数据。

(i) 变量 *favwin* 是一个二值变量，在拉斯维加斯所押的球队胜出了预定的分数差时取值 1。估计所押球队获胜概率的线性概率模型为

$$\mathrm{P}(favwin = 1 \mid spread) = \beta_0 + \beta_1 spread$$

如果分数差包括了所有相关的信息，会预期 $\beta_0 = 0.5$。解释其原因。

(ii) 用 OLS 估计第 (i) 部分的模型。相对于双侧对立假设检验 $\mathrm{H}_0: \beta_0 = 0.5$，同时使用通常的标准误和异方差稳健的标准误。

(iii) *spread* 在统计上显著吗？当 $spread = 10$ 时，被押球队获胜的估计概率是多少？

(iv) 现在对 $\mathrm{P}(favwin = 1 \mid spread)$ 估计一个 probit 模型。解释和检验截距项为零的虚拟假设。[提示：注意 $\Phi(0) = 0.5$。]

(v) 利用 probit 模型估计当 $spread = 10$ 时被押球队获胜的概率，并与第

(iii)部分的 LPM 估计值相比较。

568 (vi) 在 probit 模型中增加变量 *favhome*，*fav*25 和 *und*25，并用似然比检验来检验这些变量的联合显著性。(χ 平方分布中的自由度是多少?)解释这个结果，注意分数差是否包括了赛前可观测到的全部信息这个问题。

17.9 本题利用 LOANAPP.RAW 中的数据；也可参见习题 7.16。

(i) 估计一个 *approve* 对 *white* 的 probit 模型。求出白人和黑人贷款许可的估计概率。与线性概率估计值相比如何?

(ii) 现在在这个 probit 模型中增加变量 *hrat*，*obrat*，*loanprc*，*unem*，*male*，*married*，*dep*，*sch*，*cosign*，*chist*，*pubrec*，*mortlat*1，*mortlat*2 和 *vr*。有对非白人歧视的统计上显著的证据吗?

(iii) 用 logit 估计模型的第(ii)部分，将 *white* 的系数与 probit 估计值相比较。

(iv) 你将如何在 probit 和 logit 之间比较歧视效应的大小?

17.10 本题利用 FRINGE.RAW 中的数据。

(i) 样本中有多大百分比的工人 *pension* 等于零? 对于养老金不等于零的工人，*pension* 的取值范围为多大? 为什么 Tobit 模型适合于建立 *pension* 的模型?

(ii) 估计一个用 *exper*，*age*，*tenure*，*educe*，*depends*，*married*，*white* 和 *male* 解释 *pension* 的 Tobit 模型。白人和男性的养老金统计上显著地高些吗?

(iii) 对于同样 35 岁、单身无赡养负担、受 16 年教育和有 10 年工作经验的一个白人男子和一个非白人女子，利用第(ii)部分中的结果估计其期望养老金的差异。

(iv) 在这个 Tobit 模型中添加 *union*，并评论其显著性。

(v) 以养老金—收益比 *peratio* 作为因变量，再做第(iv)部分中的 Tobit 模型。(注意这个比值介于 0～1 之间，但常常取值为零，而永远也不会接近于 1。因此用 Tobit 模型作为一个近似很好。)性别或种族对养老金—收益比有影响吗?

17.11 在例 9.1 中，我们在 *narr*86 的一个线性模型中添加了二次项 $pcnv^2$，$ptime86^2$ 和 $inc86^2$。

(i) 在例 17.3 中，利用 CRIME1.RAW 中的数据，并将这些项添加到泊松回归中。

(ii) 计算由 $\hat{\sigma}^2=(n-k-1)^{-1}\sum_{i=1}^{n}\hat{u}_i^2/\hat{y}_i$ 给出的 $\hat{\sigma}^2$ 的估计值。有过度分散的证据吗? 应如何调整泊松 MLE 标准误?

(iii) 利用第(i)和(ii)部分中的结果和表 17.3，计算检验这三个二次项联合显著性的拟似然比统计量。

569 17.12 参见第 13 章中的表 13.1。那里我们利用 FERTIL1.RAW 中的数据，估计了妇女已生育子女数 *kids* 的一个线性模型，

(i) 利用表 13.1 中同样的变量估计 *kids* 的一个泊松回归模型，解释 *y*82 的系数。

(ii) 保持其他因素不变，黑人妇女和非黑人妇女在生育性上的估计百分比差异是多少？

(iii) 求 $\hat{\sigma}$。有过度分散和分散不足的证据吗？

(iv) 计算泊松回归中的拟合值和作为 $kids_i$ 和 $\hat{kids}_i$ 之间相关系数平方的 R-平方，并与线性回归模型中的 R-平方相比较。

17.13 利用 RECID.RAW 中的数据，通过 OLS(仅用 552 个未截取的持续期间)估计例 17.4 中的模型。一般性地评论这些估计值与表 17.4 中估计值的比较。

17.14 本题利用 MROZ.RAW 中的数据。

(i) 利用在工作的 428 个妇女的数据，通过以 $exper$，$exper^2$，$nwifeinc$，age，$kidslt6$ 和 $kidsge6$ 为解释变量的 OLS 来估计受教育的回报。报告 $educ$ 的估计值及标准误。

(ii) 现在用 Hecikt 估计受教育的回报，其中所有的外生变量在第二阶段的回归中都会出现。换句话说，就是 $\log(wage)$ 对 $educ$，$exper$，$exper^2$，$nwifeinc$，age，$kidslt6$，$kidsge6$ 和 $\hat{\lambda}$ 的回归。将估计的受教育回报及其标准误与第(i)部分的结果相比。

(iii) 只用 428 个工作妇女的观测，将 $\hat{\lambda}$ 对 $educ$，$exper$，$exper^2$，$nwifeinc$，age，$kidslt6$ 和 $kidsge6$ 回归。R-平方为多大？这如何有助于解释你在第(ii)部分得到的结果？[提示：考虑多重共线性。]

附录17A

限值因变量模型中的渐近标准误

对本章所介绍模型和方法的渐近标准误的推导，远远超出了本书的研究范围。这些推导不仅要求矩阵代数，还要求非线性估计的高级渐近理论。伍德里奇(Wooldridge, 1999)给出仔细分析这些方法和几个推导所要求的背景知识。

看一下能至少推出几种方法的渐近标准误的公式是有启发意义的。给定二值响应模型 $\mathrm{P}(y=1|\boldsymbol{x})=G(\boldsymbol{x\beta})$，其中 $G(\cdot)$是 logit 或 probit 函数，而 $\boldsymbol{\beta}$ 是 $k\times 1$ 的参数向量，$\hat{\boldsymbol{\beta}}$ 的渐近方差矩阵可估计为一个 $k\times k$ 矩阵

570

$$\mathrm{A\hat{V}ar}(\hat{\boldsymbol{\beta}}) \equiv \left(\sum_{i=1}^{n} \frac{[g(\boldsymbol{x}_i C\hat{\beta})]^2 \boldsymbol{x}'_i \boldsymbol{x}_i}{G(\boldsymbol{x}_i\hat{\boldsymbol{\beta}})[1-G(\boldsymbol{x}_i\hat{\boldsymbol{\beta}})]}\right)^{-1} \tag{17.50}$$

(对矩阵代数的简介可参见附录 D。)去掉涉及 $g(\cdot)$和 $G(\cdot)$的项，这个公式看起来很像 OLS 估计量的估计方差矩阵减掉 $\hat{\sigma}^2$。式(17.50)中的表达式解释了响应概率的非线性性质[即 $G(\cdot)$的非线性性质]和二值响应模型中异方差

性的特殊形式：$\mathrm{Var}(y|\boldsymbol{x})=G(\boldsymbol{x\beta})[1-G(\boldsymbol{x\beta})]$。

式(17.50)中对角元素的平方根就是 $\hat{\beta}_j$ 的渐近标准误，在支持 logit 和 probit 分析的计量经济软件包中都会例行地把它们报告出来。一旦我们有了这些，就能以通常的方式得到(渐近)t 统计量和置信区间。

式(17.50)中的矩阵也是对 $\boldsymbol{\beta}$ 的多元约束进行瓦尔德检验的基础［参见 Wooldridge(1999，第 15 章)］。

Tobit 的渐近方差矩阵更为复杂，但结构类似。注意到我们也可以得到 $\hat{\sigma}$ 的一个标准误。在式(17.32)中容许 $\sigma^2\neq1$，泊松回归的渐近方差具有与式(17.50)类似的结构：

$$\mathrm{A}\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}})\equiv\hat{\sigma}^2\Big(\sum_{i=1}^{n}\exp(\boldsymbol{x}_i\hat{\boldsymbol{\beta}})\boldsymbol{x}_i'\boldsymbol{x}_i\Big)^{-1}$$

这个矩阵对角元素的平方根就是渐近标准误。如果泊松假定成立，我们可以从这个公式中去掉 $\hat{\sigma}^2$(因为 $\sigma^2=1$)。

截取回归、断尾回归和 Heckit 样本选择纠正模型的渐近标准误都更复杂，尽管它们具有与前面的公式相同的特征。具体情况，可参见 Wooldridge (1999)。

第 18 章　时间序列的深入讨论

571 在本章，我们将讨论时间序列计量经济学中一些更深入的问题。在第10章、第11章和第12章，我们在几个地方强调过，由于很多经济的时间序列有趋势、持续的特点，在回归分析中使用时间序列数据时一定要小心。在本章，我们除了要讨论无限分布滞后模型及预测等方面的问题以外，还要介绍在分析有单位根的时间序列过程方面的一些最新进展。

在18.1节，我们描述了一类特殊的无限分布滞后模型，在这些模型中自变量的变化会影响因变量所有将来的值。从概念上来说，这些模型只不过是第10章中的有限分布滞后模型的扩展；但这些模型的估计还是有些挑战性的。

在18.2节，我们阐述了怎样规范地检验时间序列过程中的单位根。回想一下第11章，为了应用通常的渐近理论，我们排除了过程中有单位根的可能情况。因为单位根的出现意味着现在的一个冲击会有长久不衰的影响，所以判断一个过程是否有单位根是很重要的。

在18.3节，我们讨论了两个有单位根的时间序列过程之间的谬误回归的概念。主要的结论是，即使两个单位根序列是相互独立的，在做其中一个对另一个的回归时，还是很有可能得到一个统计上显著的 t 统计量。这个结论提醒我们，当应变量和自变量都是自积或协积的概念适用于两个过程是I(1)，但二者的线性组合却是I(0) 的情况；在这种条件下，做一个对另一个

的回归就不是谬误回归，这个回归反而告诉了我们关于它们之间长期关系的信息。两个序列之间的协积还意味着一种特殊的短期动态模型，称做误差纠正模型。18.4 节介绍了这些模型。

在 18.5 节，我们综述了预测的问题，把本章和前面几章中的所有工具都归纳起来，说明了如何使用回归方法来预测时间序列将来的结果。有关预测的文献非常之多，我们只能把注意力集中在最常见的以回归为基础的方法上。我们也谈到了葛兰杰（Granger）因果关系这一话题。

18.1 无限分布滞后模型

572 令$\{(y_t, z_t): t = \cdots, -2, -1, 0, 1, 2, \cdots\}$代表一个双变量时间序列过程（我们只观察到了它的一部分）。表示 y_t 和当期及所有过去的 z_t 之间关系的一个**无限分布滞后模型**［infinite distributed lag（IDL）model］是

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \cdots + u_t \tag{18.1}$$

式中，z 的滞后可以一直追溯到无限的过去。这个模型只是对现实的一个近似，因为没有哪个经济的过程从无限远的过去开始。与有限分布滞后模型不同的是，IDL 模型不要求在某个特定时刻截断滞后。

为了让式（18.1）有意义，随着 j 趋于无穷大，β_j 滞后的系数必须趋于零。这并不是说 δ_2 在数量上比 δ_1 小；它的意思是，z_{t-j}对 y_t 的影响必须随着 j 变的很大而最终变得很小。在大多数实际应用中，它也有相应的经济含义：很远的过去的 z 对 y 的解释能力不如较近的过去的 z。

即使我们认为式（18.1）是一个有用的模型，我们显然也不能不加限制地估计它。一个原因是，我们只能观察到数据的有限的历史。方程（18.1）涉及无限个参数，δ_0，δ_1，δ_2，…，它们不可能毫无限制地被估计出来。下面，对 δ_j 作出一些使我们得以估计式(18.1)的限制条件。

与有限分布滞后模型一样，式（18.1）的短期倾向就是 δ_0（见第 10 章）。一般来讲，δ_h 的含义与 FDL 中的系数相同。假使：$s<0$ 时，$z_s = 0$，$z_0 = 1$；$s>0$ 时，$z_s = 0$。也就是说，z 在时间 $t = 0$ 暂时增长一个单位，然后又回到它的初始值零。对所有 $h \geqslant 0$，我们都有 $y_h = \alpha + \delta_h + u_h$，所以有

$$\mathrm{E}(y_h) = \alpha + \delta_h \tag{18.2}$$

在这里，我们使用了标准的假定，即 u_t 有零均值。可见，给定 z 在时间零的一个单位的暂时变化，δ_h 就是$\mathrm{E}(y_h)$的改变值。我们刚才说过了，为了让 IDL 有意义，h 必须随着 h 变大而趋于零。这意味着，z 的一个暂时变化对 y 的期望没有长期影响：随着$h \rightarrow \infty$，$\mathrm{E}(y_h) = \alpha + \delta_h \rightarrow \alpha$。

为了便于阐述，我们假定过程 z 从 $z_s = 0$ 开始，而一个单位的增长发生在 $t = 0$。更一般地，如果 z 在时间 t 暂时地增长一个单位（相对于任意的一

个初始值)，那么，δ_h 就度量了 h 个时期后 y 的期望值的变化。滞后分布 δ_h，作为 h 的函数，表示了在给定 z 的一个单位的暂时增加量时，y 所遵循的期望路径。

模型(18.1)中的长期倾向等于所有的滞后系数之和：

$$\text{LRP}=\delta_0+\delta_1+\delta_2+\delta_3+\cdots \tag{18.3}$$

在这里，我们假定了无限和是已经被很好地定义了的。因为 δ_j 必须收敛于零，对于足够大的 p，LRP 常常能够用 $\delta_0+\delta_1+\cdots+\delta_p$ 形式的有限和很好地近似表示。573 为了理解 LRP，设当 $s<0$ 时，过程 z_t 稳定在 $z_s=0$。在时间 $t=0$，这个过程永久地增长一个单位。比如说，z_t 如果是货币供给的百分比变化，而 y_t 是通货膨胀率。那么，当 $s<0$ 时，z_s 用 0 替换；$t\geqslant0$ 时，z_t 用 1 替换，得到

$$y_h=\alpha+\delta_0+\delta_1+\cdots+\delta_h+u_h$$

式中，$h\geqslant0$ 是任何时间范围。因对任何 t，u_t 的均值均为零，所以有

$$\mathrm{E}(y_h)=\alpha+\delta_0+\delta_1+\cdots+\delta_h \tag{18.4}$$

[不妨比较一下式（18.4）和式（18.2）。] 随着预测范围的扩大，也就是，随着 h 趋于无穷，根据定义，式（18.4）的右边就是长期倾向。所以，LRP 度量了给定 z 的一个单位的永久性的增长，y 的期望值的长期变化。

问题 18.1 ☞

假使 $s<0$ 时，$z_s=0$，$z_0=1$，$z_1=1$；$s>1$ 时，$z_s=0$。请求出 $\mathrm{E}(y_{-1})$，$\mathrm{E}(y_0)$ 和 $h\geqslant1$ 时的 $\mathrm{E}(y_h)$。当 h 趋于无穷大时，情况又会如何？

前面对 LRP 的推导及 δ_j 的解释，都利用了误差有零均值这一事实；和以前一样，如果模型中包括了截距，这就不是一个很好的假定。更仔细地看看我们的推理过程，就会发现，我们假定了任何时期 z 的变化都不会对 u_t 的期望值有影响。这就是无限分布滞后条件下的严格外生性假定。我们最初在第 10 章介绍了严格外生性假定（特别是假定 TS.2)。无限分布滞后条件下的严格外生性假定可以正式地表述为

$$\mathrm{E}(u_t|\cdots,z_{t-2},z_{t-1},z_t,z_{t+1},\cdots)=0 \tag{18.5}$$

它使得 u_t 的期望值不依赖于任何时期的 z。虽然式（18.5）在很多应用中是很自然的，但它也排除了其他很重要的可能情况。实际上，式(18.5)不允许 y_t 对将来的 z 的反作用，因为在 $h>0$ 时，z_{t+h} 必须与 u_t 不相关。在通货膨胀/货币供给增长的例子中，y_t 是通货膨胀率，z_t 是货币供给的增长，式(18.5)排除了由于今天的通货膨胀率而造成的将来的货币估计的增长这种情况。因为货币供给政策常常是以利率和通货膨胀保持在一定水平上为目标的，所以上面的假定可能不太合理。

我们将在下一节介绍到估计 δ_j 的一个方法，要以严格外生性假定为条件，才能得到 δ_j 的一致的估计值。更弱一点的假定是

$$\mathrm{E}(u_t \mid z_t, z_{t-1}, \cdots) = 0 \tag{18.6}$$

在式(18.6)下，误差与现在的和过去的 z 都不相关，但它有可能与将来的 z 相关；这使得 z_t 服从于一种政策规则，这个规则取决于过去的 y。在一些情况下，式(18.6)对估计 δ_j 来说就足够了；我们将在下一小节解释这方面原因。

574 要提醒你注意的是，式(18.5)和式(18.6)都没有提到 $\{u_t\}$ 在序列相关方面的性质。（这与有限分布滞后模型中的情形相同。）我们认为 $\{u_t\}$ 应该是序列相关的，因为式(18.1)在11.4节的意义上一般不是动态完整的。我们将在后面研究序列相关的问题。

如果式(18.6)成立而式(18.5)不成立，我们如何来解释滞后系数和LRP呢？答案很简单：和以前一样。我们仍然可以做一个想像（或非事实）的实验，尽管我们观测到的数据是由 y_t 和将来的 z 之间的相互作用产生的。例如，我们当然可以研究货币供给增长的一个永久性增加量对通货膨胀有什么影响，尽管货币供给的增长不能用严格外生性来描述。

几何（或考依克）分布滞后

因为一般都有无数个 δ_j，我们不可能毫无约束地一致地估计出它们。式(18.1)最简单的一种形式是**几何（或考依克）分布滞后**（geometric or Koyck distributed lag，GDL），虽然它仍取决于无限个滞后，但在这个模型中，每个 δ_j 都只取决于两个参数：

$$\delta_j = \gamma\rho^j, \ |\rho| < 1, \ j = 0, 1, 2, \cdots \tag{18.7}$$

参数 γ 和 ρ 可以是正的也可以是负的，但 ρ 的绝对值一定要小于1。这保证了随着 $t \to \infty$，δ_j 趋于零。实际上，这种收敛的速度很快。（例如，当 $\rho = 0.5$，$j = 10$ 时，$\rho^j = 1/1024 < 0.001$。）

GDL的即期倾向(IP)是 $\delta_0 = \gamma$，所以IP的符号由 γ 的符号决定。比如说，如果 $\gamma > 0$，且 $\rho > 0$，那么所有的滞后系数都是正的。如果 $\rho < 0$，滞后系数的符号则相反（j 为奇数时，ρ^j 是负的）。长期倾向就比较难求，不过，我们可以使用下面这个几何序列之和的标准结果：对于 $|\rho| < 1$，$1 + \rho + \rho^2 + \cdots + \rho_j + \cdots = 1/(1-\rho)$，所以

$$\mathrm{LRP} = \gamma/(1-\rho)$$

LRP与 γ 的符号相同。

如果我们把式(18.7)代入到式(18.1)中，会得到一个模型，它依赖于无限远的过去的 z。但是，只要我们做个简单的减法运算，就可以得到一个可以估计的模型。把时间 t 和 $t-1$ 的IDL写出来：

$$y_t = \alpha + \gamma z_t + \gamma\rho z_{t-1} + \gamma\rho^2 z_{t-2} + \cdots + u_t \tag{18.8}$$

和

$$y_{t-1} = \alpha + \gamma z_{t-1} + \gamma\rho z_{t-2} + \gamma\rho^2 z_{t-3} + \cdots + u_{t-1} \tag{18.9}$$

如果把第二个方程两边乘以 ρ，然后从第一个方程减去，就只剩下了几项：

$$y_t-\rho y_{t-1}=(1-\rho)\alpha+\gamma z_t+u_t-\rho u_{t-1}$$

我们可以把它写为

575 $$y_t=\alpha_0+\gamma z_t+\rho y_{t-1}+u_t-\rho u_{t-1} \tag{18.10}$$

式中，$\alpha_0=(1-\rho)\alpha$。这个方程看起来像一个有滞后应变量的标准方程，而且 z_t 同时出现在方程中。因为 γ 是 z_t 的系数，而 ρ 是 y_{t-1} 的系数，我们可以估计出这些参数。[如果由于某些原因，我们想知道 α，估计完 γ 和 α_0 后总能求出 $\hat{\alpha}=\hat{\alpha}_0/(1-\hat{\rho})$。]

式（18.10）过于简单了，容易导致错误结论。因为这个方程中的误差 $u_t-\rho u_{t-1}$ 一般会与 y_{t-1} 相关。从式（18.9）中可以清楚地看出，u_{t-1} 和 y_{t-1} 是相关的。因此，我们把方程（18.10）写为

$$y_t=\alpha_0+\gamma z_t+\rho y_{t-1}+v_t \tag{18.11}$$

式中，$v_t\equiv u_t-\rho u_{t-1}$，所以，$v_t$ 和 y_{t-1} 一般都是相关的。如果没有其他假定，(18.11) 的 OLS 估计会得到 γ 和 ρ 的不一致的估计值。

当 u_t 独立于 z_t 和 z 和 y 的所有过去值的时候，就会出现 v_t 和 y_{t-1} 必定相关的情况。于是，式（18.8）就是动态完整的，而且，u_t 与 y_{t-1} 不相关。根据式（18.9），v_t 和 y_{t-1} 之间的协方差是 $-\rho\mathrm{Var}(u_{t-1})=-\rho\sigma_u^2$，它只有在 $\rho=0$ 时才为零。我们很容易就可以看出来，v_t 是序列相关的：因为 $\{u_t\}$ 是序列不相关的，所以有 $\mathrm{E}(v_tv_{t-1})=\mathrm{E}(u_tu_{t-1})-\rho\mathrm{E}(u_{t-1}^2)-\rho\mathrm{E}(u_tu_{t-2})+\rho^2\mathrm{E}(u_{t-1}u_{t-2})=-\rho\sigma_u^2$。当 $j>1$ 时，$\mathrm{E}(v_tv_{t-j})=0$。所以，$\{v_t\}$ 是一个一阶移动平均过程（见 11.1 节）。以上内容给出了一个派生模型的例子，它是从我们最初关心的那个模型派生出来的，有一个滞后的应变量和某种特定类型的序列相关。

如果我们作出严格外生性假定（18.5），那么 z_t 就与 u_t 和 u_{t-1} 都不相关，并因此也与 v_t 不相关。于是，如果能够为 y_{t-1} 找到合适的工具变量，我们就可以用 IV 法估计式（18.11）。怎样选择 y_{t-1} 的好 IV 呢？根据假定，u_t 和 u_{t-1} 都与 z_{t-1} 不相关，而且 v_t 也因此与 z_{t-1} 不相关。如果 $\gamma\neq0$，即使除掉 z_t 的局部影响，z_{t-1} 和 y_{t-1} 也是相关的。因此，我们可以用工具（z_t，z_{t-1}）去估计式（18.11）。一般来讲，需要根据 $\{v_t\}$ 中的序列相关来校正标准误，具体校正的方法我们在 15.7 节中已经讲过了。

代替 IV 的另一种办法利用了 $\{u_t\}$ 可能包含某种特定形式的序列相关这一事实。具体地说，除了式（18.6）以外，假定 $\{u_t\}$ 遵循 AR（1）模型

$$u_t=\rho u_{t-1}+e_t \tag{18.12}$$

$$\mathrm{E}(e_t|z_t,y_{t-1},z_{t-1},\cdots)=0 \tag{18.13}$$

很重要的是，出现在式（18.12）中的 ρ 与式（18.11）中 y_{t-1} 所乘的参数是相同的。如果式（18.12）和式（18.13）成立，可以得出

$$y_t = \alpha_0 + \gamma z_t + \rho y_{t-1} + e_t \tag{18.14}$$

576 在式(18.13)成立时，它是一个动态完整的模型。由第11章知，可以用OLS求出这些参数的一致的、渐近正态的估计量。这就很方便了，不需要去解决误差的序列相关的问题。如果 e_t 满足同方差假定 $\mathrm{Var}(e_t \mid z_t, y_{t-1}) = \sigma_e^2$，就可以采用通常的推断方法。一旦我们估计出了 γ 和 ρ，就能够很容易地估计LRP了：$\widehat{\mathrm{LRP}} = \hat{\gamma}/(1-\hat{\rho})$。

上面的方法比较简单，这完全决定于潜在的强假定，即 $\{u_t\}$ 遵循一个有与式（18.7）中相同的参数的一个AR(1)过程。这个假定通常不劣于假定 $\{u_t\}$ 是序列不相关的。不管怎样，因为估计量的一致性严重依赖于这个假定，我们最好先检验一下它。一个简单的检验方法是，先设定 $\{u_t\}$ 为一个AR(1)过程，但它有不一样的参数，比如 $u_t = \lambda u_{t-1} + e_t$。麦克莱恩和伍德里奇（McClain and Wooldridge，1995）设计了一种简单的拉格朗日乘数方法来检验 H_0：$\lambda = \rho$，可以在用OLS估计方程（18.14）后计算出它。

几何分布滞后模型可以扩展到多元解释变量的情形——每个解释变量都有无限的DL——但是，那样的话，我们必须可以把 $z_{t-j,h}$ 的系数写作 $\gamma_h \rho^j$。换句话说，虽然对每个解释变量来说 γ_h 是不同的，但 ρ 却是一样的。于是，我们可以写出

$$y_t = \alpha_0 + \gamma_1 z_{t1} + \cdots + \gamma_k z_{tk} + \rho y_{t-1} + v_t \tag{18.15}$$

在只有一个 z 的情况下出现的问题，也会出现在有多个 z 的情况下。在对式（18.12）和式（18.13）作自然的延伸——用 $\mathbf{z}_t =（z_{t1}，\cdots，z_{tk}）$ 来代替 z_t——时，OLS是一致和渐近有效的。或者，也可以使用IV方法。

有理分布滞后模型

几何DL意味着一个相当受限制的滞后分布。当 $\gamma > 0$ 且 $\rho > 0$ 时，δ_j 是正的，而且单调地下降到零。要得到更一般的无限分布滞后模型是可能的。GDL是所谓**有理分布滞后模型**［rational distributed lag（RDL）model］的一种特例。一般性的处理方法超出了我们的研究范围——哈维（Harvey，1990）给出比较好的阐述——我们只介绍一种简单实用的拓广方法。

这种RDL模型常常通过在方程（18.11）中加进一期滞后得到

$$y_t = \alpha_0 + \gamma_0 z_t + \rho y_{t-1} + \gamma_1 z_{t-1} + v_t \tag{18.16}$$

式中，$v_t = u_t - \rho u_{t-1}$，这与以前是一样的。通过反复迭代可以证明，式（18.16）与无限分布滞后模型是等价的：

$$\begin{aligned} y_t &= \alpha + \gamma_0(z_t + \rho z_{t-1} + \rho^2 z_{t-2} + \cdots) \\ &\quad + \gamma_1(z_{t-1} + \rho z_{t-2} + \rho^2 z_{t-3} + \cdots) + u_t \\ &= \alpha + \gamma_0 z_t + (\rho\gamma_0 + \gamma_1)\ z_{t-1} + \rho(\rho\gamma_0 + \gamma_1)\ z_{t-2} \\ &\quad + \rho^2(\rho\gamma_0 + \gamma_1)\ z_{t-3} + \cdots + u_t \end{aligned}$$

在这里，我们仍然需要假定$|\rho|<1$。从最后一个等式，我们可以看出滞后分布的情况。特别是，即期倾向为γ_0，而在$h>1$时，z_{t-h}的系数是$\rho^{h-1}(\rho\gamma_0+\gamma_1)$。因此，这个模型允许即期倾向的符号与其他的滞后系数不同，即使$\rho>0$。然而，如果$\rho>0$，那么，对于所有的$h\geqslant1$，δ_h就与$(\rho\gamma_0+\gamma_1)$有相同的符号。图18.1画出了$\rho=0.5$，$\gamma_0=-1$，$\gamma_1=1$时的滞后分布。

计算长期倾向的最简单的办法是，把y和z在所有的t置于某个长期值，比如说y^*和z^*上，然后我们就可以求出y^*相对于z^*的变化（见习题10.3）。我们有$y^*=\alpha_0+\gamma_0z^*+\rho y^*+\gamma_1z^*$，解出它的解为$y^*=\alpha_0/(1-\rho)+(\gamma_0+\gamma_1)/(1-\rho)z^*$。接下来，我们利用$\mathrm{LRP}=\Delta y^*/\Delta z^*$得到

$$\mathrm{LRP}=(\gamma_0+\gamma_1)/(1-\rho)$$

因为$|\rho|<1$，LRP的符号由$\gamma_0+\gamma_1$决定，而且，当且仅当$\gamma_0+\gamma_1=0$时LRP为零，如图18.1所示。

577

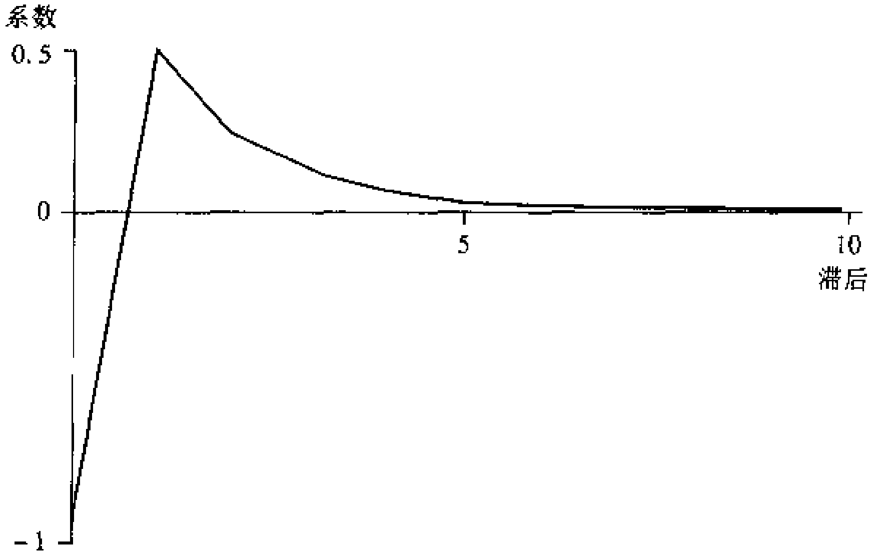


图18.1　$\rho=0.5$，$\gamma_0=-1$，$\gamma_1=1$时，有理分布滞后（18.16）的滞后分布

例18.1　房产投资和住宅价格膨胀

我们将OLS分别应用于式（18.14）和式（18.16），来估计基本的几何分布滞后模型和有理分布滞后模型。应变量是除掉了线性趋势以后的log(*invpc*)［就是说，我们对log(*invpc*)做了线性地除趋势］。z_t用于表示价格指数的上升。我们就可以估计住宅价格膨胀是如何影响房产投资围绕其趋势而波动的了。利用HSEINV.RAW中的数据，估计出的结果列于表18.1中。

由于$gprice_{-1}$非常显著，几何分布滞后模型明显被数据否定了。校正的R-平方也表明RDL拟合得更好。

这两个模型给出的长期倾向的估计结果差别非常大。如果我们错误地使用了GDL，估计到的LRP接近5：住宅价格通胀率的一个1%的永久性上涨使房产投资增长4.7%（超出它的趋势值）。从经济的角度来看，这是不可能

578 **表 18.1**　　**房产投资的分布滞后模型**

应变量：log（*invpc*），除了趋势的		
自变量	几何 DL	理性 DL
gprice	3.108	3.256
	(0.933)	(0.970)
y_{-1}	0.340	0.547
	(0.132)	(0.152)
$gprice_{-1}$	—	−2.936
		(0.973)
常数项	−0.001	−0.578
	(0.018)	(0.307)
长期倾向	4.688	0.706
样本容量	41	40
校正的 *R*-平方	0.375	0.504

的。通过估计有理分布滞后模型得到的 LRP 的值小于 1。实际上，我们在任何显著性水平上都不能拒绝虚拟假设 H_0：$\gamma_0+\gamma_1=0$（p 值 $=0.83$），所以没有证据表明 LRP 异于零。这是个很好的例子，它说明了，因为遗漏相关滞后而错误设定模型的动态是如何导致错误的结论的。

18.2　单位根的检验

我们下面来研究单位根的检验问题。在第 11 章，我们介绍了一些不够清楚的、不正式的原则，来判断一个序列是否是I(1)。在很多情况下，需要规范地检验单位根。但是，我们在后面将会知道，一定要慎重使用这些检验方法。

579 检验单位根的一种最简单的方法是从AR(1)模型开始的：

$$y_t=\alpha_0+\rho y_{t-1}+e_t,\quad t=1,2,\cdots \tag{18.17}$$

式中，y_0 为观测到的初始值。在这一节，我们都令 $\{e_t\}$ 代表在给定 y 的过去值时均值为零的一个过程：

$$\mathrm{E}(e_t\mid y_{t-1},y_{t-2},\cdots,y_0)=0 \tag{18.18}$$

[在式（18.18）下，$\{e_t\}$ 被称做对 $\{y_{t-1},\ y_{t-2},\ \cdots\}$ 的**鞅差分序列**（martingale difference sequence）。如果假定 $\{e_t\}$ 是均值为零的 *i.i.d.*，且独立于 y_0，那么它也满足式（18.18）。]

如果 $\{y_t\}$ 遵循式（18.17），当且仅当 $\rho=1$ 时它有单位根。如果 $\alpha=0$，

且 $\rho=1$，$\{y_t\}$ 就遵循一个没有漂移的随机游走［要新生值 e_t 满足(18.18)］。如果 $\alpha\neq0$，且 $\rho=1$，$\{y_t\}$ 是有漂移的随机游走，于是，$E(y_t)$ 是 t 的线性函数。有漂移的随机游走的表现与没有漂移的随机游走的表现非常不同。尽管如此，人们经常在虚拟假设中留下一个 α 不予以设定，这种做法是很常见的，我们就采用这种方法。于是，虚拟假设是，$\{y_t\}$ 有一个单位根：

$$H_0:\rho=1 \tag{18.19}$$

几乎在所有的情况下，我们只对单侧对立检验感兴趣：

$$H_1:\rho<1 \tag{18.20}$$

（实际上，它意味着 $0<\rho<1$，因为当 $\rho<0$ 时，我们所怀疑的序列有单位根的可能性极小。对立假设 H_1：$\rho>1$ 很少被认可，因为它意味着 y_t 是发散的。实际上，如果 $\alpha>0$，在 $\rho>1$ 时，y_t 的均值就有一个指数上升趋势。

当 $|\rho|<1$ 时，$\{y_t\}$ 是一个平稳的AR(1)过程，也就是说，它是弱相依或渐近有效的。回想一下第 11 章的内容，当 $|\rho|<1$ 时，有 $\mathrm{Corr}(y_t, y_{t+h})=\rho^h\to0$。因此，检验模型（18.17）中的式(18.19)，以式(18.20)为对立面，实际上是检验 $\{y_t\}$ 是I(1)，还是I(0)。［我们不把虚拟假设设为I(0)的原因是，对于严格在 -1～1 之间的 ρ 的任何值，$\{y_t\}$ 都是I(0)，而这种情况是用经典假设检验所不容易解决的。当然，也有把I(0)设为虚拟假设而把I(1)设为对立假设的一些检验办法，但要采取不同的思路。可以参考诸如 Kwiatkowski，Phillips，Schmidt 和 Shin（1992）。］

进行单位根检验的一种简便的方法是，在方程（18.17）两边同时减去 y_{t-1}，并设 $\theta=\rho-1$：

$$\Delta y_t=\alpha+\theta y_{t-1}+e_t \tag{18.21}$$

在式（18.18）下，这是个动态完整的模型，于是相对于 H_1：$\theta<0$ 检验 H_0：$\theta=0$ 就是理所当然的了。问题在于，在 H_0 成立时，y_{t-1} 是I(1)，所以通常的中心极限定理就不适用了，而 t 统计量的渐近标准正态分布是以这个定理为前提的：即使在大样本的条件下，t 统计量也不再有渐近标准正态分布。t 580 统计量在 H_0 下的渐近分布被称为**迪基-富勒分布**（Dickey-Fuller distribution），它是以迪基（Dickey）和富勒（Fuller）的名字命名的。

虽然不能使用通常的临界值，我们却可以使用式（18.21）中 θ 的通常的 t 统计量，至少在相应的临界值被列出来以后我们可以这样做。以这种思路为基础的方法被称为单位根的**迪基-富勒检验**［Dickey-Fuller（DF）test］。用来计算渐近临界值的理论实在有些复杂，在高级的时间序列计量经济学的教材里有相关的介绍。［参见 Banerjee，Dolado，Galbraith 和 Hendry（1993），BDGH，等等。］不过，使用这些理论的结果却是件容易的事。t 统计量的临界值已经被几位学者列成了表格，其中，最初的工作是从迪基和富勒（Dickey and Fuller，1979）开始的。表 18.2 选自 BDGH（1993，表 4.2），它列出了不同显著性水平上的大样本的临界值。（针对小样本进行调整了的临界值也可以在 BDGH 中找到。）

表 18.2　　单位根 t 检验的渐近临界值：无时间趋势

显著性水平	1%	2.5%	5%	10%
临界值	−3.43	−3.12	−2.86	−2.57

如果 $t_\theta < c$，c 是表 18.2 中的一个负值，我们就拒绝虚拟假设 $H_0: \theta = 0$，而取 $H_1: \theta < 0$。例如，在 5% 的显著性水平上进行检验，如果 $t_\theta < -2.86$，我们就拒绝虚拟假设。这要求 t 统计量的大小要远大于（指在绝对值上——译者注）使用标准正态临界值时的大小，后者是 -1.65。如果我们用标准正态临界值来检验单位根，在 H_0 实际上为真的时候拒绝 H_0 的可能性会远大于 5%。

例 18.2　3 个月期国债利率的单位根检验

我们利用 INTQRT.RAW 中的季度数据来检验 3 个月期国债利率中的单位根。估计式（18.20），得到

$$\begin{aligned} \Delta \hat{r}3_t &= 0.625 - 0.091\, r3_{t-1} \\ &\quad (0.261)\quad(0.037) \\ &\quad n = 123,\ R^2 = 0.048 \end{aligned} \tag{18.22}$$

式中，按照惯例，我们在估计值下面的括号里报告了标准误。需要提醒你注意的是，我们不能用这些标准误来构造置信区间，也不能用它们来进行传统的 t 检验，因为有单位根存在时它们的行为和平常不一样。$r3_{t-1}$ 的系数表 581 明 ρ 的估计值为 $\hat{\rho} = 1 + \hat{\theta} = 0.909$。虽然 $\hat{\rho}$ 是小于 1 的，我们却无法判断 ρ 是否是统计上小于 1 的。$r3_{t-1}$ 的 t 统计量是 $-0.091/0.037 = -2.46$。根据表 18.2，10% 水平上的临界值是 -2.57；所以我们在 10% 的水平上无法拒绝 H_0: $\rho = 1$ 而接受 H_1: $\rho < 1$。

和其他的假设检验一样，当我们不能拒绝 H_0 时，也不能说我们就可以接受 H_0。为什么呢？假使在前面的例子中我们用标准的 t 统计量来检验 H_0: $\rho = 0.9$——这个 t 统计量是渐近有效的，因为在 H_0 下 y_t 是 I(0)。然后，我们求出 $t = 0.001/0.037$，这个值很小，不能足以否定 $\rho = 0.9$。但是，我们同时接受 $\rho = 1$ 和 $\rho = 0.9$ 是没有意义的。

当不能拒绝单位根时，就像前面的例子中那样，我们只能得出结论，认为数据没有提供足够的证据以否定 H_0。在这个例子中，所进行的检验的确提供了一定的否定 H_0 的证据，因为 t 统计量接近于 10% 的临界值。（计算 p 值是比较理想的做法，但由于非正态分布的缘故，我们需要特殊的软件。）另外，虽然 $\hat{\rho} \approx 0.91$ 意味着 $\{r3_t\}$ 有一定的持续性，但是，$\rho = 0.9$ 时的AR

(1)模型中距离为 10 个时期的观测值之间的相关性大约是 0.35，而不是 $\rho=1$ 时近似的相关性 1。

如果我们用 $r3_t$ 作回归分析中的解释变量，情况会怎样呢？单位根检验的结果意味着我们应该非常慎重：如果 $r3_t$ 有单位根，通常的渐近逼近法不一定成立（我们在第 11 章讨论过）。一个解决的方法是，在任何分析中都使用 $r3_t$ 的一阶差分。在 18.4 节我们将了解到，它并不是惟一的办法。

在有更复杂的动态的模型中，我们也需要检验单位根。如果 y_t 遵循 $\rho=1$ 的式（18.17），那么 Δy_t 就是序列不相关的。通过用更多的滞后来扩展方程（18.21），很容易就可以使得$\{\Delta y_t\}$遵循一个 AR 模型。比如：

$$\Delta y_t=\alpha+\theta y_{t-1}+\gamma_1\Delta y_{t-1}+e_t \tag{18.23}$$

式中，$|\gamma_1|<1$。这个条件确保了在 H_0：$\theta=0$ 下，y_t 遵循一个平稳的AR(1)模型。在对立假设 H_1：$\theta<0$ 条件下，可以证明$\{y_t\}$遵循一个平稳的AR(2)模型。

更一般地说，我们可以在方程中加进 Δy_t 的 p 期滞后，用以表示这个过程的动态。单位根检验的虚拟假设的检验方法大体相似——做下面的回归：

$$\Delta y_t\ 对\ y_{t-1},\Delta y_{t-1},\cdots,\Delta y_{t-p} \tag{18.24}$$

并对 y_{t-1}的系数 $\hat{\theta}$ 进行 t 检验，方法与从前一样。这种扩展了的迪基-富勒检验常常被称为**扩展的迪基-富勒检验**（augmented Dickey-Fuller test，DF），因为这个回归通过增加滞后的变化 Δy_{t-h}得到了扩展。临界值和拒绝的规则与以前是一样的。在式（18.24）中增添滞后变化的目的在于消除 Δy_t 中的 582 序列相关。在式（18.24）中加进的滞后越多，我们丢掉的初始观测值也越多。如果添加的滞后太多，小样本的性质就会使检验的功效大打折扣。但是，如果加进的滞后太少，检验的大小（指显著水平——译者注）就是不正确的，即使在渐近的意义上。因为表 18.2 中的临界值的有效性是以模型设定的动态完整性为前提的，滞后的长度往往由数据的周期（及样本大小）决定。对于年度数据，一个或两个滞后就够了。对于月度数据，模型中可以包含 12 个滞后。但是，无论在什么情况下，都没有固定的规则。

有意思的是，滞后改变量的 t 统计量有近似的 t 分布。任何一组 Δy_{t-h} 的联合分布的 F 统计量也是渐近有效的。（这些都以 11.5 节讨论的同方差假定为条件。）因此，我们可以利用标准的检验判断在式（18.24）中是否有足够多的滞后改变量。

例 18.3　美国年度通货膨胀的单位根检验

利用以消费者价格指数CPI为基础的美国年度通货膨胀数据，我们检验一下通货膨胀中的单位根（见 PHILLIPS.RAW）。时间从 1948 年到 1996 年。在扩展的迪基-富勒回归中加入 Δinf_t 的一期滞后，得到

$$\begin{aligned}\Delta\widehat{inf}_t&=1.36-0.310inf_{t-1}+0.138\Delta inf_{t-1}\\&\quad\ (0.261)(0.103)\qquad(0.126)\end{aligned}$$

$$n=47,\ R^2=0.172$$

单位根检验的 t 统计量是 $-0.310/0.103=-3.01$。因为5%的临界值是 -2.86，我们在5%的水平上拒绝存在单位根的虚拟假设。ρ 的估计值大约是0.690。总的来看，有足够强的证据可以否定通货膨胀中存在单位根。滞后 Δinf_{t-1} 有一个约为1.10的 t 统计量，所以我们没有必要把它包含在模型中，但我们不可能事先预见到这一点。如果我们去掉 Δinf_{t-1}，反对单位根的证据就变得更强一点：$\hat{\theta}=-0.335$（$\hat{\rho}=0.665$）且 $t_{\hat{\theta}}=-3.13$。

对于有明显趋势的时间序列，我们需要对这个单位根检验进行修改。一个趋势—平稳过程——它的均值有线性趋势，但在趋势上又是I(0)的——可能会被错误地当做单位根过程，如果我们不在迪基-富勒回归中加进一个时间趋势的话。换句话说，如果我们对一个有趋势但又是I(0)的过程做通常的DF或扩展的DF检验时，拒绝单位根的功效可能会很小。

为了分析有时间趋势的序列，我们把基本的方程变为

$$\Delta y_t=\alpha+\delta t+\theta y_{t-1}+e_t \tag{18.25}$$

式中，虚拟假设仍然是 H_0：$\theta=0$，对立假设是 H_1：$\theta<0$。在对立假设下，y_t 是一个趋势—平稳过程。如果 $\{y_t\}$ 有单位根，那么就有 $\Delta y_t=\alpha+\delta t+e_t$，所以除非 $\delta=0$，否则 y_t 的改变量的均值对 t 为线性。[可以证明，$\mathrm{E}(y_t)$ 是 t 的二次方。]经济的序列在一阶差分后有一个线性趋势是很不正常的，所以一个更为合适的虚拟假设也许是 H_0：$\theta=0$，$\delta=0$。虽然用 F 检验来检验联
583 合假设是有可能的——但对临界值要进行修正——用 t 检验来仅仅检验 H_0：$\theta=0$ 是更常见的做法。我们在这里就采用这种方法。[更多关于联合检验的内容参见BDGH（1993，4.4节）。]

当我们在模型中加进时间趋势时，这个检验的临界值就变了。直观地看，这是因为对一个单位根过程除趋势，会使得它更像一个I(0)过程。所以，我们要求 t 统计量是更大的值，才能拒绝 H_0。包含时间趋势的 t 检验的迪基-富勒临界值被列在表18.3中，这个表取自BDGH（1993，表4.2）。

表18.3　单位根的 t 检验的渐近临界值：线性时间趋势

显著性水平	1%	2.5%	5%	10%
临界值	−3.96	−3.66	−3.41	3.12

例如，为了在5%的水平上拒绝单位根假设，我们需要 θ 的 t 统计量小于 -3.41，在没有时间趋势时这个数据应该是 -2.86。

与在没有趋势的情况一样，我们可以在方程（18.25）中补充进 Δy_t 的滞后，以预防序列相关。这并不会影响到我们怎样进行检验。

例 18.4　美国实际国内生产总值对数中的单位根

我们可以把有时间趋势的单位根检验应用于 INVEN.RAW 中美国的 GDP 数据。这些年度数据是从 1959 年到 1995 年的。我们检验 log（GDP_t）中是否有单位根。这个序列被认为是有趋势的，而这个趋势看起来大致是线性的。我们在模型中加进 $\Delta\log$（GDP_t）的一期滞后，$\Delta\log(GDP_t)$就是（以小数形式表示的）GDP 的增长（*growth*），从而表示出模型的动态：

$$\begin{aligned} g\hat{GDP}_t &= 1.65 + 0.005\,9t - 0.210\log(GDP_{t-1}) + 0.264gGDP_{t-1} \\ &\quad (0.67)(0.002\,7)\ (0.087) \qquad\qquad\quad (0.165) \qquad (18.26) \\ &\quad n = 35,\ R^2 = 0.268 \end{aligned}$$

通过这个方程，得到 $\hat{\rho} = 1 - 0.21 = 0.79$，它明显小于 1。但我们不能拒绝在 GDP 的 log 中的单位根：$\log(GDP_{t-1})$的 t 统计量是 $-0.210/0.087 = -2.41$，它比 10% 水平的临界值 -3.12 大很多。$gGDP_{t-1}$的 t 统计量是 1.60，它在 10% 的水平上的双侧检验上差不多是显著的。

关于单位根，有什么结论呢？还是不能拒绝单位根，尽管 ρ 的点估计不是特别接近 1。当我们有一个较小的样本容量时——$n = 35$ 被认为是个很小的
584 数值——如果过程中有接近于单位根的某些东西，拒绝单位根是件困难的事。利用更长时期的更多的数据，很多研究人员发现几乎没有反对 $\log(GDP)$有单位根假设的证据。这使得他们中的很多人都假定 GDP 的增长是I(0)，也就是说，log（GDP）是I(1)。不过，由于现在所能获得的样本容量较小，我们对这种观点还不太相信。

如果忽略掉时间趋势，反对 H_0 的证据就少得多了，因为 $\hat{\theta} = -0.023$且 $t_{\hat{\theta}} = -1.92$。这一回，ρ 的估计值更接近于 1 了，但由于忽略了时间趋势，得出的结论是错误的。

你可能会有这样的想法：把式（18.26）中的时间趋势的 t 统计量与标准正态分布或 t 分布的临界值进行比较，来看看时间趋势是否是显著的。不幸的是，趋势的 t 统计量没有渐近的标准正态分布（除非 $|\rho| < 1$）。这个 t 统计量的渐近分布是知道的，但很少用到它。一般而言，我们靠直觉（或时间序列的图形）来判断是否应该在 DF 检验中加进一个趋势。

还有很多种单位根的检验方法。其中的一种是把截距从回归中省略掉，它只适用于明显没有趋势的序列，也就是，式（18.21）中的 α 被设为零。这种迪基-富勒检验的变型很少被用到，因为如果 $\alpha \neq 0$ 就会有偏误。此外，我们也可以考虑更为复杂的时间趋势，比如二次的时间趋势。不过，这类方法也很少用到。

另一类检验方法试图用不同的方式来控制 Δy_t 中的序列相关，而不是在式（18.21）或式（18.25）中加入滞后项，其方法与我们在 12.5 节讨论过的 OLS 估计量的序列相关—稳健标准误有关。主要思想是，对 Δy_t 中的序列相关尽可能表明无所知之。在实际计算中，（扩展的）迪基-富勒检验已经

表现得很不错了。［关于其他检验方法的讨论，参见 BDGH（1993，4.3节）。］

18.3 谬误回归

在横截面环境下，我们用“谬误回归”这个词来描述这样一种情形，也就是，两个变量通过分别与第三个变量的相关关系建立起联系。但是，当我们控制了第三个变量，比如说 z 时，x 对 y 的偏影响变成了零。很自然，这种情况也会发生在I(0)变量的时间序列分析中。

正如我们在 10.5 节中讨论过的，在有上升或下降的趋势时间序列之间可能会找到一种谬误的关系。若这些序列关于它们的时间趋势是弱相依的，则只要在回归模型中加进一个时间趋势，就可以很好地解决问题了。

我们在处理一阶单积过程时，会遇到新的麻烦。做两个独立I(1)序列的回归，即使两个各自的均值没有趋势，也常常会得到一个显著的 t 统计量。

更具体地看，令 $\{x_t\}$ 和 $\{y_t\}$ 是由下面的方程产生的随机游走：

$$x_t = x_{t-1} + a_t \tag{10.27}$$

和

585

$$y_t = y_{t-1} + e_t, \quad t = 1, 2, \cdots \tag{18.28}$$

式中，$\{a_t\}$ 和 $\{e_t\}$ 都是独立同分布的新生值，它们的均值为零，方差分别为 σ_a^2 和 σ_e^2。为了简单起见，将初始值设为 $x_0 = y_0 = 0$。还假定 $\{a_t\}$ 和 $\{e_t\}$ 是相互独立的过程。这意味着 $\{x_t\}$ 和 $\{y_t\}$ 也是相互独立的。但是，如果我们做下面的回归：

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t \tag{18.29}$$

并求出 $\hat{\beta}_1$ 的通常的 t 统计量和通常的 R-平方，情况会是怎样的呢？因为 y_t 和 x_t 之间是独立的，我们希望 $\text{plim}\hat{\beta}_1 = 0$。更为重要的是，如果我们在 5% 的水平上检验 H_0：$\beta_1 = 0$ 及其对立的 H_1：$\beta_1 \neq 0$，我们希望 $\hat{\beta}_1$ 的 t 统计量在每 100 次有 95 次（也就是 95% 的时候）都是不显著的。通过模拟，葛兰杰和纽博尔德（Granger and Newbold，1974）证明了事实并非如此：即使 y_t 和 x_t 之间是独立的，在很大比例的次数里，y_t 对 x_t 的回归都会产生一个统计上显著的 t 统计量，而且，远大于名义上的显著性水平。葛兰杰和纽博尔德把这种现象称为**谬误回归问题**（spurious regression problem）：y 和 x 之间根本没有关系，但用了 t 统计量的 OLS 回归往往标示它们之间存在某种关系。

最新的模拟由戴维森和麦金农（Davidson and MacKinnon，1993，表 19.1）给出，在他们的模拟中，a_t 和 e_t 都由独立同分布正态随机变量产生，总共产生了10 000个样本。当样本容量 $n = 50$ 时，在 5% 的置信水平上，H_0：$\beta_1 = 0$ 的双侧检验中的标准 t 统计量在大约 66.2% 的次数都拒绝 H_0；而不

只是在 5% 的时间里拒绝 H_0。随着样本容量的增大，情况变得更糟了：当 $n=250$时，虚拟假设在 84.7% 的次数都被拒绝。

让我们来看看如果做水平的 y 对水平的 x 的回归会出现什么情况。把式（18.27）的背景模型写为

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{18.30}$$

为了使 $\hat{\beta}_1$ 的 t 统计量在大样本的情况下有一个近似的标准正态分布，至少 $\{u_t\}$ 应该是一个零均值且序列不相关的过程。但在 H_0：$\beta_1=0$ 下，$y_t=\beta_0+u_t$，因为 y_t 是一个初始值 $y_0=0$ 的随机游走，所以仅当 $\beta_0=0$，而且更重要的是，仅当 $u_t=y_t=\sum_{j=1}^{t} e_j$ 时，方程（18.30）才在 H_0 的条件下成立。也就是说，u_t 在 H_0 下是个随机游走。这显然连第 11 章的高斯-马尔科夫假定的渐近意义都失掉了。

问题 18.2

如果 $\{x_t\}$ 和 $\{y_t\}$ 分别由式（18.27）和式（18.28）产生，且 $\{a_t\}$ 和 $\{e_t\}$ 都是 *i.i.d.* 序列。做 Δy_t 对 Δx_t 的回归，那么，斜率系数（比如 $\hat{\gamma}_1$）的 *plim* 是什么？$\hat{\gamma}_1$ 的 t 统计量有怎样的行为？

在方程中添加一个时间趋势，也不会改变我们的结论。如果 y_t 或 x_t 是
586 一个有漂移的随机游走，而我们却没有在模型中包含时间趋势，谬误回归的问题将更加严重。如果 $\{a_t\}$ 和 $\{e_t\}$ 是一般的 I(0) 过程，而不是 *i.i.d.* 过程，也有同样性质的结论。

通常的 t 统计量除了不以标准正态分布为其极限外——实际上，随着 $n\to\infty$，它会增长到无限大——R-平方的行为也是非标准的。在横截面条件下或在有 I(0) 时间序列变量的回归中，R-平方依概率收敛于总体 R-平方：$1-\sigma_u^2/\sigma_y^2$。但在 I(1) 过程的谬误回归中不会有这样的事情。R-平方并没有一个明确的 *plim*，它实际上收敛于一个随机变量。把这一思想正式表述出来，超出了本书的研究范围。[对 t 统计量和 R-平方渐近性质的讨论可以在 BDGH（3.1 节）中找到。] 通过以上内容得到的结论是，假使 y_t 和 x_t 是相互独立的时间序列，R-平方是个很大的数值的概率很高。

在多个独立变量都是 I(1)，或者其中的一些是 I(0) 的时，也存在同样的问题。如果 y_t 是 I(1)，而且，至少有一些解释变量是 I(1)，所得的回归结果就可能是谬误的。

有 I(1) 变量的回归可能是谬误回归，这个重要的问题，促使经济学家们重新审视那些 t 统计量非常显著和 R 平方特别高的总量宏观经济变量的时间序列回归。在下一节，我们将证明，只有当 I(1) 应变量和某个 I(1) 自变量确实有关系时，做它们的回归才能提供有价值的信息。

18.4 协积和误差纠正机制

前一节讨论的谬误回归问题，会使人们对回归分析中的水平I(1)变量担心。在更前一些的章节中，我们建议，在把I(1)变量用在线性回归模型中之前，无论是用OLS还是用工具变量法来估计它们，都应该取它们的差分。这当然是个很稳妥的办法。在葛兰杰、纽博尔德最初关于谬误回归的论文发表之后，很多时间序列回归都采用了这种方法。不过，总是对I(1)变量取差分，使我们可以回答的问题的范围受到了限制。

协　积

协积（cointegration）的概念，由恩格尔和葛兰杰（Engle and Granger，1987）正式提出，这个概念使得涉及I(1)变量的回归也可能有意义。完整地分析协积问题需要很多数学知识，我们在这里只介绍一些基本的问题和在实际应用中需要的方法。

如果$\{y_t: t=0, 1, \cdots\}$和$\{x_t: t=0, 1, \cdots\}$是两个I(1)过程，那么，一般而言，对于任何β，$y_t - \beta x_t$都是一个I(1)过程。但是，对某些$\beta \neq 0$的值来说，$y_t - \beta x_t$有可能是个I(0)过程——这意味着，它有常数均值、固定的方差及只与任意两个变量之间的时间间隔有关的自相关，而且，它是渐近不相关的。如果这样的β存在，我们就说y和x是协积的，并称β为协积系数。[当然，我们也可以考虑$\gamma \neq 0$时的$x_t - \gamma y_t$：因为如果$y_t - \beta x_t$是I(0)，那么，$x_t - (1/\beta)y_t$也是I(0)。所以，y_t和x_t的线性组合不是惟一的，但如果我们把y_t的系数固定为1，它们的线性组合就是惟一的。见习题18.3。为简单起见，我们只考虑$y_t - \beta x_t$形式的线性组合。]

为了便于说明，我们不妨取$\beta = 1$，假令$y_0 = x_0 = 0$，并写$y_t = y_{t-1} + r_t$，587 $x_t = x_{t-1} + v_t$，其中的$\{r_t\}$和$\{v_t\}$是两个I(0)过程，它们的均值都为零。于是，y_t和x_t都有到处流浪，不会经常地回归到原地（初始值零）的倾向。与此相反，如果$y_t - x_t$是I(0)，它就有零均值，并且会有规律地向零回归。

下面举个具体的例子。令$r6_t$代表6个月国债（在季度t末）的年利率，$r3_t$代表3个月国债的年利率。（这些都被称为债券等价收益，在金融栏目里有关于它们的报告。）在例18.2中，利用INTQRT.RAW的数据，我们没有发现证据能否定假设：$r3_t$有单位根；$r6_t$也如此。定义6个月和3个月国债利率的差价为$spr_t = r6_t - r3_t$。那么，利用方程（18.21），得到spr_t的迪基-富勒t统计量是-7.71（$\hat{\theta} = -0.67$或$\hat{\rho} = 0.33$）。因此，我们有很强的理由拒绝spr_t中有单位根这一假设，所以认为spr_t是I(0)。结论是，尽管$r6_t$和$r3_t$都是单位根过程，但它们之间的差却是I(0)过程。换句话说，

$r6_t$ 和 $r3_t$ 是协积的。

问题 18.3 令 $\{(y_t, x_t): t = 1, 2, \cdots\}$ 是二元时间序列，其中每个都是没有漂移的I(1)。如果 y_t 和 x_t 是协积的，那么 y_t 和 x_{t-1} 也是协积的，请解释为什么。

这个例子中的协积，同很多例子中的一样，有其经济含义。如果 $r6_t$ 和 $r3_t$ 不是协积的，这两种利率之间的差异可能会变得很大，没有走到一起的倾向。但是，这是不可能的。假使差价 spr_t 在几个时期都持续增大，使得6个月国债成为一种更理想的投资品种。那么，投资者就会从3个月国债转向6个月国债，从而使得6个月国债的价格上升，3个月国债的价格下降。因为利率是与价格反方向变动的，$r6$ 就会下降，$r3$ 则上升，直到差价消失为止。所以，$r6$ 和 $r3$ 之间比较大的差距不可能持续下去：差价有向其均值靠拢的倾向。（这个差价实际上有很小的正的均值，因为长期投资者相对于短期投资者获得的回报更高些。）

差价 spr_t 不会在较长时期都偏离其平均值这一事实，可以用另一种方法来描述：$r6$ 和 $r3$ 有一种长期的关系。为了说明我们的意思，令 $\mu = \mathrm{E}(spr_t)$ 代表差价的期望值。于是，可以写

$$r6_t = r3_t + \mu + e_t$$

式中，$\{e_t\}$ 为一个均值为零的I(0)过程。当 $e_t = 0$ 或 $r6^* = r3^* + \mu$ 时，均衡或长期关系就会出现。在任何时期，都可能发生对均衡的偏离，但这种偏离是暂时的：存在着经济的力量把 $r6$ 和 $r3$ 往均衡关系上拉。

在利率的例子中，我们利用经济利率指出，如果 y_t 和 x_t 是协积的，β
588 的值应该是多少？如果我们有一个 β 的假设值，那么检验两个序列是不是协积的就很简单了：只要定义一个新的变量 $s_t = y_t - \beta x_t$，然后用前面讲过的通常的DF和扩展的DF方法来检验 $\{s_t\}$ 就行了。如果我们拒绝 $\{s_t\}$ 中存在单位根的假设，而倾向于I(0)这个假设，那么，y_t 和 x_t 就是协积的。也就是说，虚拟假设是，y_t 和 x_t 不是协积的。

当（可能的）协积系数 β 未知时，对协积的检验就更难了。在检验 $\{s_t\}$ 中的单位根之前，我们必须首先估计 β。如果 y_t 和 x_t 是协积的，从下面回归中得出的OLS估计量 $\hat{\beta}$

$$y_t = \hat{\alpha} + \hat{\beta}x_t \tag{18.31}$$

与 β 是一致的。问题在于，虚拟假设声称这两个序列不是协积的，这就意味着，在 H_0 条件下，我们做了一个谬误回归。幸运的是，即使 β 是估计得到的，我们也有可能把临界值列成表。在估计 β 时，我们对来自于式（18.31）中的残差，比如说 $\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$，运用DF或扩展DF检验。惟一的区别是，这里的临界值考虑到了对 β 的估计方法。表18.4给出了渐近临界值。它们来自于DM（1993，表20.2）。

表 18.4　　　　协积检验的渐近临界值：无时间趋势

显著性水平	1%	2.5%	5%	10%
临界值	3.90	−3.59	−3.34	3.04

在基本的检验中，我们做 $\Delta\hat{u}_t$ 对 $\Delta\hat{u}_{t-1}$ 的回归，然后将 $\Delta\hat{u}_{t-1}$ 的 t 统计量与表 18.4 中的临界值进行比较。如果 t 统计量小于临界值，我们认为 $y_t-\beta x_t$ 对于某个 β 是I(0)，也就是，y_t 和 x_t 是协积的。我们可以在模型中加进 $\Delta\hat{u}_t$ 的滞后，以防止序列相关。如果我们比较一下表 18.4 中的临界值和表 18.2 中的临界值，就会发现，为了证明协积的存在，我们必须得到一个比通常的 DF 临界值在绝对值上更大的 t 统计量。这是因为，OLS 把残差平方和最小化了，它倾向于产生出看起来像I(0)序列的残差，即使在 y_t 和 x_t 没有协积关系时也是这样。

如果 y_t 和 x_t 不是协积的，y_t 对 x_t 的回归就是谬误的，不能告诉我们什么有价值的信息：因为 y 和 x 之间没有任何长期关系。我们仍可以做一阶差分形式的变量 Δy_t 和 Δx_t，及其滞后的回归。但是在解释这些回归时要适当：它们用 x 的差分解释了 y 的差分，与水平的关系就没有必然联系。

如果 y_t 和 x_t 是协积的，就可以利用这种协积关系来设定更为一般的动态模型，在下一小节你将看到这方面的内容。

前面的讨论假定了无论 y_t 还是 x_t 都没有漂移。这个假定对于利率来说是合适的，但对其他的时间序列就未必合适。如果 y_t 和 x_t 包含漂移项，$\mathrm{E}(y_t)$ 和 $\mathrm{E}(x_t)$ 就是时间的（通常是增）函数。协积的严格的定义要求 $y_t-\beta x_t$ 是没有趋势的I(0)。为了看看都需要些什么条件，我们写出 $y_t=\delta t+g_t$ 和 $x_t=$ 589
$\lambda t+h_t$，其中的 $\{g_t\}$ 和 $\{h_t\}$ 是I(1)过程，δ 是 y_t 中的漂移 $[\delta=\mathrm{E}(\Delta y_t)]$，而 λ 是 x_t 中的漂移 $[\lambda=\mathrm{E}(\Delta x_t)]$。现在，如果 y_t 和 x_t 是协积的，必定存在一个 β，使得 $g_t-\beta h_t$ 是I(0)。但是，

$$y_t-\beta x_t=(\delta-\beta\lambda)t+(g_t-\beta h_t)$$

一般是个趋势—平稳过程。协积的严格形式要求没有趋势，所以有 $\delta=\beta\lambda$。对于有漂移的I(1)过程，随机的部分——也就是 g_t 和 h_t——有可能是协积的，但导致 $g_t-\beta h_t$ 是I(0)的参数 β 没有消除这个线性时间趋势。

我们可以检验 g_t 和 h_t 之间的协积，而不用考虑趋势部分，只要做回归：

$$\hat{y}_t=\hat{\alpha}+\hat{\eta}t+\hat{\beta}x_t \tag{18.32}$$

并对残差 $\Delta\hat{u}_t$ 做通常的 DF 或扩展的 DF 检验。渐近的临界值被列在表 18.5 中。[来自于 Davidson 和 MacKinnon(1993，表 20.2)。]

在这种情况下，协积的存在允许 $y_t-\beta x_t$ 有线性趋势的可能性，但至少表明它不是I(1)。

表 18.5　　协积检验的渐近临界值：线性时间趋势

显著性水平	1%	2.5%	5%	10%
临界值	4.32	4.03	-3.78	3.50

例 18.5　生育率和个人税收豁免之间的协积

在第 10 章和第 11 章，我们用各种各样的模型估计了美国的总生育率（gfr）和个人税收豁免（pe）的实际金额之间的关系。水平形式和差分形式的静态回归结果有很大的差异。水平形式的回归，在包含时间趋势时，给出 pe 的 OLS 系数等于 0.187（se = 0.035），而 $R^2 = 0.500$。在一阶差分形式下（没有趋势），Δpe 的系数是 -0.043（se = 0.028），而 $R^2 = 0.032$。尽管还有其他的原因造成了这些差异——比如分布滞后动态被错误设定——水平和差分形式的回归之间的差别提醒我们要检验协积。当然，这需要假定 gfr 和 pe 是 I(1) 过程。看来情况的确如此：在扩展的 DF 检验中，有一个滞后的改变量和一个线性时间趋势，它们都有大约为 -1.47 的 t 统计量，而且估计到的那些 ρ 都接近于 1。

如果我们做 gfr 对 t 和 pe 的回归，求出残差，并应用有一个滞后的扩展的 DF 检验，得到 $\Delta \hat{u}_{t-1}$ 的 t 统计量为 -2.43，它离 10% 的临界值 -3.50 比较远。因此，我们的结论是，几乎没有证据表明 gfr 和 pe 之间有协积，即使考虑到它们都有各自的趋势的可能情况，结论也是相同的。很可能的是，我们在前面得到的水平形式的回归结果受到了谬误回归问题的牵连。590

令人欣慰的是，当我们使用一阶差分形式并添加两个滞后时——见方程（11.27）——我们发现了 Δpe 对 Δgfr 有大体上正的、显著的长期影响。

如果我们认为两个序列是协积的，就常常想要检验有关协积系数的假设。例如，一种理论可能会认为协积系数为 1。比较理想的情况是，我们可以用一个 t 统计量来检验这个假设。

下面，我们来介绍无时间趋势的情况，把它延伸到有线性趋势的情况是容易的。当 y_t 和 x_t 是 I(1) 并且有协积关系时，可以写

$$y_1 = \alpha + \beta x_t + u_t \tag{18.33}$$

式中，u_t 为均值为零的 I(0) 过程。一般来讲，$\{u_t\}$ 中包含了序列相关，但通过第 11 章我们知道，这并不影响 OLS 的一致性。我们以前提到过，把 OLS 用于式（18.33）可以一致地估计出 β（或 α）。糟糕的是，因为 x_t 是 I(1) 的，通常的推断程序未必适用：OLS 不是渐近正态分布的，而且 $\hat{\beta}$ 的 t 统计量不一定有渐近的 t 分布。我们从第 10 章了解到，如果 $\{x_t\}$ 是严格外生的——见假定 TS.2——而且误差是同方差、序列不相关和正态分布的，

OLS估计量也是正态分布的（以解释变量为条件），而且 t 统计量刚好是 t 分布的。不过，这些条件在大多数场合显得过于苛刻了。而协积的概念根本没有涉及$\{x_t\}$和$\{u_t\}$之间的关系问题，除了要求 u_t 是I(0)以外，也没有限制 u_t 中的序列相依性。

幸运的是，式（18.33）的特性，即$\{x_t\}$没有严格外生性，虽然使得推断变得非常困难，但并非无法解决。因为 x_t 是I(1)，相应的严格外生性的概念是，对于所有的 s 和 t，u_t 与 x_s 不相关。在 s 与 t 比较接近时，通过把 u_t 写成 Δx_s 的函数，得到一组新的误差，这时，就可以使严格外生性假定得到满足，至少是近似地满足。比如

$$\begin{aligned} u_t = \eta + \phi_0\Delta x_t + \phi_1\Delta x_{t-1} + \phi_2\Delta x_{t-2} \\ + \gamma_1\Delta x_{t+1} + \gamma_2\Delta x_{t+2} + e_t \end{aligned} \tag{18.34}$$

式中，通过构造，e_t 与出现在方程中的每个 Δx_s 都不相关。我们希望 e_t 与 Δx_s 的其他的滞后和超前值都不相关。我们知道，随着$|s-t|$变大，e_t 和 Δx_s 之间的相关性趋于零，因为它们是I(0)过程。现在，我们把式（18.34）代入式（18.33），得到

$$\begin{aligned} y_t = \alpha_0 + \beta x_t + \phi_0\Delta x_t + \phi_1\Delta x_{t-1} + \phi_2\Delta x_{t-2} \\ + \gamma_1\Delta x_{t+1} + \gamma_2\Delta x_{t+2} + e_t \end{aligned} \tag{18.35}$$

这个方程看起来有点怪，将来的 Δx_s 与当前值 x_t 及 Δx_t 的滞后一起出现在方程中。问题的关键在于，x_t 的系数仍然是β，而且，通过构造，x_t 现在是严格外生的了。严格外生性假定是求得 $\hat{\beta}$ 的渐近正态 t 统计量的一个重要条件。

591 从式（18.35）中得出的β的OLS估计值被称做β的**超前和滞后估计量**（leads and lags estimator），这是因为它利用Δx 的方法而得此名。[参见 Stock 和 Watson（1993）等。] 在式（18.35）中惟一令我们担心的是，$\{e_t\}$中存在序列相关的可能性。这个问题可以通过计算 $\hat{\beta}$ 的序列相关—稳健标准误（12.5节有相关的内容）或使用标准的AR(1)纠正（比如 Cochrane-Orcutt 方法）来解决。

例 18.6　利率的协积参数

在前面，我们检验了3个月和6个月国债利率之间的协积关系——假定了协积参数为1。这使得我们找到了协积，并很自然地得出结论：协积参数等于1。尽管如此，我们还是来直接估计一下协积参数并检验 H_0：$\beta=1$。我们采用超前和滞后估计值，用$\Delta r3$的两个超前和两个滞后，及它当期的改变量来求它。β的估计值是$\hat{\beta}=1.038$，通常的OLS标准误是0.008 1。所以，H_0：$\beta-1$ 的 t 统计量是$(1.038-1)/.008\ 1\approx4.69$，它是拒绝 H_0 的有力证据。（当然，1.038在经济上是否异于1也是个需要考虑的问题。）因为没有什么证据表明残差中有序列相关，我们就可以把 t 统计量当做有渐近的正态分布。[与此不同的是，没有$\Delta r3$这些项时——因而可多利用4次观测

——β 的 OLS 估计值是1.026(se=0.007 7)。但得自式（18.33）的 t 统计量不一定是有效的。]

有很多种其他的协积参数估计量，这一直是比较活跃的研究领域。协积的概念可以应用于多于两个的过程中，但相应的解释、检验和估计都更为复杂。另外还有一个问题，即使我们把一个系数规范化为1，还可能同时存在着很多种协积关系。BDGH 提供了这方面的讨论和一些文献。

误差纠正模型

协积的概念不但使我们了解了两个序列之间潜在的长期关系，也大大丰富了我们可以处理的动态模型的范围。如果 y_t 和 x_t 是I(1)过程且不是协积的，我们可能会估计差分形式的一个动态模型。举个例子，考虑方程

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + u_t \tag{18.36}$$

式中，给定 Δx_t，Δy_{t-1}，Δx_{t-1}及其他的滞后；u_t 的均值为零。这个方程实质上就是方程（18.16），只不过是一阶差分形式而不是水平形式。如果把它当做一个有理分布滞后模型，我们可以求出即期倾向、长期倾向和 Δy 的、用 Δx 的分布滞后来表达的滞后分布。

如果 y_t 和 x_t 是协积的，协积参数为 β，我们就又有了一个I(0)变量，592 于是我们可以把它放在式（18.36）中。令 $s_t = y_t - \beta x_t$，这里的 s_t 是I(0)，而且，为简单起见，我们假定 s_t 有零均值。现在，我们可以把 s_t 的滞后加到方程中去了。在最简单的情况下，我们只加进 s_t 的一个滞后：

$$\begin{aligned}\Delta y_t &= \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \delta s_{t-1} + u_t \\ &= \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + u_t\end{aligned} \tag{18.37}$$

式中，$\mathrm{E}(u_t | I_{t-1}) = 0$；$I_{t-1}$包含关于 Δx_t 和所有 x 和 y 的过去值的信息。$\delta(y_{t-1} - \beta x_{t-1})$ 这一项被称为**误差纠正项**（error correction term），而式（18.37）是**误差纠正模型**（error correction model）的一个例子。（在一些误差纠正模型中，x 的同期变化 Δx_t 被省略了。是否把它包括进去部分地取决于方程的目的。在预测中，Δx_t 极少被包括进去，具体的原因我们将在18.5节谈到。）

误差纠正模型使得我们可以研究 y 和 x 之间关系的短期动态。看一个简单的例子，考虑一个没有 Δy_t 和 Δx_t 的滞后模型：

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + \delta(y_{t-1} - \beta x_{t-1}) + u \tag{18.38}$$

式中，$\delta<0$。如果 $y_{t-1} > \beta x_{t-1}$，那么前一个时期的 y 已经超过了均衡水平；因为 $\delta<0$，误差纠正项会把 y 往回拉，使它回到均衡水平。类似地，如果

$y_{t-1}<\beta x_{t-1}$，误差纠正项就会使 y 朝着向均衡返回的方向有一个正的变化。

如何来估计误差纠正模型中的那些参数呢？如果我们知道 β，就很容易了。例如，在式（18.38）中，我们只要做 Δy_t 对 Δx_t 和 s_{t-1} 的回归就行了，其中的 $s_{t-1}=(y_{t-1}-\beta x_{t-1})$。

例 18.7 持有收益率的误差纠正模型

在习题 11.6 中我们做了 $hy6_t$ 对 $hy3_{t-1}$ 的回归，其中 $hy6_t$ 是在时间 $t-1$ 买入 6 个月国债并在时间 t 把它当做 3 个月国债卖出时获得的 3 个月的持有收益率（百分数），$hy3_{t-1}$ 为在时间 $t-1$ 买入 3 个月国债而获得的 3 个月持有收益率。期望值假设意味着回归的斜率系数不会在统计上异于 1。但是，有迹象表明 $\{hy3_t\}$ 中有单位根，这使得我们对标准的回归分析无法信赖。我们将假定这两个持有收益率都是 I(1) 过程。期望值假设至少表示 $hy6_t$ 和 $hy3_{t-1}$ 有协积关系且 β 等于 1（见习题 18.14）。在这个假定下，误差纠正模型为

$$\Delta hy6_t=\alpha_0+\gamma_0\Delta hy3_{t-1}+\delta(hy6_{t-1}-hy3_{t-2})+u_t$$

式中，给定时间为 $t-1$ 或更早的 $hy3$ 和 $hy6$ 的值；u_t 有零均值。误差纠正模型中怎样设变量的滞后，根据期望值假设来裁定。

利用 INTQRT.RAW 中的数据，得到

$$\begin{aligned}\Delta hy6_t &= \underset{(0.043)}{0.090}+\underset{(0.264)}{1.218}\Delta hy3_{t-1}-\underset{(0.244)}{0.840}(hy6_{t-1}-hy3_{t-2}) \\ n &= 122,\ R^2=0.790\end{aligned} \qquad (18.39)$$

误差纠正系数是负的，并且很显著。举例来说，如果 6 个月国债的持有收益
593 率比 3 个月高出 1%，$hy6$ 在下一个季度平均就下降 0.84%。有意思的是，$\hat{\delta}=-0.84$，它不是统计上异于 -1，只要计算 95% 的置信区间就可以看出这一点。

问题 18.4 在持有收益的误差纠正模型中，你如何检验 H_0：$\gamma_0=1$，$\delta=-1$？

在很多其他的例子中，协积参数必须估计出来。那么，我们把 s_{t-1} 换成 $\hat{s}_{t-1}=y_{t-1}-\hat{\beta}x_{t-1}$，其中的 $\hat{\beta}$ 可以是各种 β 的估计量。我们已经学过了 OLS 估计量以及超前和滞后估计量，但这时又出现了一个问题：$\hat{\beta}$ 的抽样变异会怎样影响误差纠正模型中对其他参数的推断呢？幸而，根据恩格尔和葛兰杰（Engle and Granger，1987），我们可以（渐近地）忽略初步估计 β 所造成的影响。这就很方便了。用 $\hat{\beta}$ 代替 β 的方法被称为**恩格尔-葛兰杰两步法**（Engle-Granger two-step procedure）。

18.5 预测

在一些经济学的分支中，预测经济的时间序列是相当重要的，它现在仍是一个很活跃的研究领域。在本节，我们集中研究以回归为基础的预测方法。戴博尔德（Diebold，1998）对预测方面的问题作了全面的介绍，其中包括最新的进展情况。

在这一节，假定我们主要的精力放在预测一个时间序列过程的将来值上面，而不是要估计因果性或结构性经济模型。

首先，我们要介绍一些与模型的具体形式无关的预测方面的基本理论。假设我们在时间 t 想要预测 y 在时间 $t+1$ 的结果，即 y_{t+1}。所用时期单位可以是一年、一个季度、一个月、一个星期或一天。令 I_t 代表我们在时间 t 可以观测到的所有信息。这个**信息集**（information set）包含 y_t 及 y 的以前的值，还常常包括其他变量的时间为 t 及更早的值。我们可以用无数种方法来组合这些信息从而预测 y_{t+1}。有没有哪种方法是最好的呢？

答案是肯定的，只要设定我们允许的由预测误差所带来的损失。令 f_t 代表在时间 t 所做的对 y_{t+1} 的预测。我们称 f_t 为**超前一步预测**（one-step-ahead forecast）。**预测误差**（forecast error）是 $e_{t+1}=y_{t+1}-f_t$，y_{t+1} 的结果一旦被观测到了，它的值就知道了。度量损失最常见的是误差的平方，多元线性回归模型的最小二乘估计也是根据它推出来的。对预测误差的平方来说，正的和负的误差是等价的，越大的误差得到的权重越大。例如，+2 和 −2 带来的损失是一样的，它们都是预测误差为 +1 和 −1 时的损失的 4 倍。预测误差的平方是**损失函数**（loss function）的一种。另一种常见的损失函数是预测误差的绝对值 $|e_{t+1}|$。由于下面将要谈到的原因，我们现在集中研究误差平方损失。

决定了采用误差平方损失函数，我们就可以决定怎样来最大限度地利用 594 时间 t 的信息来预测 y_{t+1} 了。但我们必须认识到，在时间 t，我们并不知道 e_{t+1}，因为 y_{t+1} 是个随机变量，e_{t+1} 也就是个随机变量。所以，任何一种选择 f_t 的标准必须以我们在时间 t 知道的信息为基础。很自然的想法是，当 I_t 给定时，就选使得预测误差平方的期望最小的预测值：

$$\mathrm{E}(e_{t+1}^2 \mid I_t)=\mathrm{E}[(y_{t+1}-f_t)^2 \mid I_t] \tag{18.40}$$

概率论的一个基本结论是（见附录 B 中的性质 CE.6），条件均值 $\mathrm{E}(y_t \mid I_t)$ 使式（18.40）最小化。换句话说，给定时间 t 的信息，如果想要最小化预测误差平方的期望，预测应该是，在时间 t 我们所知道的所有变量都给定时的 y_{t+1} 的期望值。

对于很多常见的时间序列过程，这个条件期望值很容易求出来。假设 $\{y_t: t=0, 1, \cdots\}$ 是个**鞅差分序列**（martingale difference sequence，MDS），并

设 I_t 为 $\{y_t, y_{t-1}, \cdots, y_0\}$，它就是观测到的过去的 y。根据定义，对所有的 t，有 $\mathrm{E}(y_{t+1}|I_t)=0$：在时间 t 对 y_{t+1} 最好的预测永远是零！在 18.2 节我们讲过，一个均值为零的 $i.i.d.$ 序列是个鞅差分序列。

鞅差分序列中，过去对预测将来毫无用处。股票收益被广泛地认为非常接近 MDS，或许会有正的均值。无论怎样，关键是 $\mathrm{E}(y_{t+1}|y_t, y_{t-1}, \cdots)=\mathrm{E}(y_{t+1})$：条件均值等于非条件均值，在这种情况下，过去的 y 对预测将来的 y 没有任何帮助。

如果对所有的 $t \geqslant 0$ 都有 $\mathrm{E}(y_{t+1}|y_t, y_{t-1}, \cdots, y_0)=y_t$，过程 $\{y_t\}$ 就是**鞅**（martingale）。［如果 $\{y_t\}$ 是一个鞅，那么 $\{\Delta y_t\}$ 就是一个鞅差分序列，后者的名称由此而来。］下个时期 y 的预测值总是现在这个时期的 y 值。

下面是更复杂一点的模型：

$$\mathrm{E}(y_{t+1}|I_t)=\alpha y_t+\alpha(1-\alpha)y_{t-1}+\cdots+\alpha(1-\alpha)^t y_0 \qquad (18.41)$$

式中，$0<\alpha<1$ 是我们必须作出选择的参数。这种预测方法被称做**指数平滑法**（exponential smoothing），因为滞后的 y 的权重按指数率趋于零。

把期望值写成式（18.41）的理由是，由它可以得到非常简单的**递推**（recurrence）关系。设 $f_0=y_0$，于是，$t \geqslant 1$ 时，预测值可以通过下面的等式求出：

$$f_t=\alpha y_t+(1-\alpha)f_{t-1}$$

换句话说，y_{t+1} 的预测值是 y_t 和在时间 $t-1$ 对 y_t 的预测值的加权平均。指数平滑法只适用于某些特定的时间序列，而且还需要选择 α。我们下面将要介绍的回归方法，就灵活得多。

前面的讨论围绕着只提前一个时期预测 y。在时间 t 预测 y_{t+h}（h 为正整数）过程中出现的一般性的问题也基本类似。特别是，如果我们用预测误差的期望作为损失的度量，最好的预测值仍然是 $\mathrm{E}(y_{t+h}|I_t)$。在处理**超前多步预测**（multiple-step-ahead-forecast）时，我们用 $f_{t,h}$ 来表示在时间 t 对 y_{t+h} 的预测。

用于预测的各种回归模型

595 有很多种不同的回归模型可以用来预测时间序列的将来值。第 10 章中用于时间序列数据的第一个回归模型是静态模型。下面就来看看怎样用这个模型进行预测。假定我们只有一个解释变量：

$$y_t=\beta_0+\beta_1 z_t+u_t \qquad (18.42)$$

暂时假定 β_0 和 β_1 已知。在时间 $t+1$ 时，方程为 $y_{t+1}=\beta_0+\beta_1 z_{t+1}+u_{t+1}$。现在如果我们在时间 t 知道 z_{t+1}，以至于它也是 I_t 的一个组成部分，而且 $\mathrm{E}(u_{t+1}|I_t)=0$，那么，

$$\mathrm{E}(y_{t+1}|I_t)=\beta_0+\beta_1 z_{t+1}$$

式中，I_t 包括 z_{t+1}，y_t，z_t，…，y_1，z_1。方程的右边是在时间 t 所做的 y_{t+1} 的预测值。这种预测的方法常被称为**条件预测**（conditional forcast），因为它是以我们知道 z 在时间 $t+1$ 的值为条件的。

不幸的是，无论在任何时候，我们都很少能够知道将来的解释变量的值。除非是在有时间趋势和季节性变量的情况，我们将在下面介绍这方面的问题。除此之外，我们很少能在时间 t 知道 z_{t+1}。有时，我们希望能够对若干个 z_{t+1} 找出条件预测值。

作为预测模型的（18.42）还有一个问题：$\mathrm{E}(u_{t+1}|I_t)=0$ 意味着 $\{u_t\}$ 不能含有序列相关，但我们知道，在大多数静态回归模型中，这一点多少有些不真实。[习题 18.8 要求你在有AR(1)误差的简单的分布滞后模型中推导出预测值。]

如果我们在时间 t 不知道 z_{t-1}，就不能把它包括在 I_t 中。于是有

$$\mathrm{E}(y_{t+1}|I_t)=\beta_0+\beta_1\mathrm{E}(z_{t+1}|I_t)$$

这意味着，为了预测 y_{t+1}，我们必须依据相同的信息集首先预测 z_{t+1}。这种方法通常被称做**非条件预测**（unconditional forecast），因为我们没有假定在时间 t 时知道 z_{t+1}。事实上，这么称谓也不够妥当，因为我们的预测仍然是以信息 I_t 为条件的。但是，这一名称在预测文献中已经根深蒂固了。（只不过是它用在预测中时，被我们赋予一个特定的含义罢了。）

就预测而言，除非由于某些原因使我们无法避免使用式（18.42）中的静态模型，否则，最好还是设定一个只取决于 y 和 z 的滞后值的模型。这会使我们的工作量减少一些，不必在预测 y 之前还要预测右边的变量。很容易想到的一个模型是

$$y_t=\delta_0+\alpha_1 y_{t-1}+\gamma_1 z_{t-1}+u_t$$
$$\mathrm{E}(u_t|I_{t-1})=0 \tag{18.43}$$

式中，I_{t-1}包括时间为 $t-1$ 及更早的 y 和 z。于是，在时间 t 所做的 y_{t+1} 的预测值就是 $\delta_0+\alpha_1 y_t+\gamma_1 z_t$；如果我们知道这些参数的值，那么只要把 y_t 和 z_t 的值代进去就行了。

如果只想用 y 的过去值估计 y 的将来值，那么可以从式（18.43）里面去掉 z_{t-1}。也可以在方程中加进 y 或 z 的更多的滞后，或其他变量的滞后。特别是对于超前一步预测来说，这类模型非常有用。

超前一步预测

596 用像式（18.43）这样的模型来计算样本末期的下一个时期的预测值是相当简单的。令 n 代表样本容量，y_{n+1}的预测值是

$$\hat{f}_n=\hat{\delta}_0+\hat{\alpha}_1 y_n+\hat{\gamma}_1 z_n \tag{18.44}$$

在这里，假定参数已经用 OLS 估计出来了。我们在 f_n 上面加一个帽（^）

符，目的在于强调估计了回归模型中的参数。（如果我们知道这些参数，预测值就不会有估计误差了。）预测误差——我们只有到时间 $n+1$ 才能知道它的值——是

$$\hat{e}_{n+1} = y_{n+1} - \hat{f}_n \tag{18.45}$$

如果我们在预测方程中加入 y 或 z 的更多滞后，样本中的前面的几次观测值就会被丢掉更多。

y_{n+1}的预测值 $\hat{f}_n$ 常被称做**点预测**（point forecast）。我们也能够获得一个**预测区间**（forecast interval）。预测区间与我们在 6.4 节所讲的**预报区间**（prediction interval）实质上是一样的。在那里我们介绍了，在经典线性模型假定下如何求出恰好 95% 的预报区间。预测区间也采用了完全相同的计算方法。如果模型不满足经典线性模型假定——比如说，模型中包含滞后的应变量，如同式（18.44）那样——那么，只要在 I_{t-1} 给定时，u_t 是均值为零、方差恒定的正态分布（这一条件保证了 OLS 统计量是渐近正态分布的，有通常的 OLS 方差；它也使得 u_{n+1} 与均值为零、方差为 σ^2 的 OLS 估计量是相互独立的），预测区间就仍然是渐近正确的。令 se（$\hat{f}_n$）表示预测值的标准误，令 $\hat{\sigma}$ 为回归的标准误。［由 6.4 节，我们可以做 y_t 对 $(y_{t-1}-y_n)$ 和 $(z_{t-1}-z_n)$ 的回归；也就是说，在做回归之前在每个滞后的 y 中减去 y 在时间 n 的值，对 z 也做同样的处理，这样求出的截距和它的标准误就是我们所要的 $\hat{f}_n$ 和 $\mathrm{se}(\hat{f}_n)$。］于是

$$\mathrm{se}\ (\hat{e}_{n+1}) = \{[\mathrm{se}\ (\hat{f}_n)]^2 + \hat{\sigma}^2\}^{1/2} \tag{18.46}$$

而（渐近的）95% 的预测区间是

$$\hat{f}_n \pm 1.96 \cdot \mathrm{se}\ (\hat{e}_{n+1}) \tag{18.47}$$

因为 $\mathrm{se}(\hat{f}_n)$ 大致与 $1/\sqrt{n}$ 成正比，所以 $\mathrm{se}(\hat{f}_n)$ 相对于误差 u_{n+1} 中的由 $\hat{\sigma}$ 来度量的不确定性来说，往往是比较小的。［一些计量经济学软件包自动计算预测区间，另外一些则需要做些简单的手工计算来求出式（18.47）。］

例 18.8　预测美国的失业率

利用 PHILLIPS.RAW 中 1948—1996 年的数据，我们来预测一下美国公
597 民在 1997 年的失业率。我们使用两个模型。第一个是 $unem$ 的简单的 AR(1) 模型：

$$\begin{aligned} \widehat{unem}_t &= 1.572 + 0.732\,unem_{t-1} \\ &\quad\ (0.577)\quad (0.097) \\ &\quad\ n = 48,\ \bar{R}^2 = 0.544,\ \hat{\sigma} = 1.049 \end{aligned} \tag{18.48}$$

在第二个模型中，我们加进有一年滞后的通货膨胀：

$$\begin{aligned} \widehat{unem}_t &= 1.304 + 0.647\,unem_{t-1} + 0.184\,inf_{t-1} \\ &\quad\ (0.490)\quad (0.084)\qquad\quad (0.041) \\ &\quad\ n = 48,\ \bar{R}^2 = 0.677,\ \hat{\sigma} = 0.883 \end{aligned} \tag{18.49}$$

方程（18.49）中滞后的通货膨胀率非常显著（$t\approx4.5$），而且从第二个方程得到的校正的 R-平方比第一个中的高很多。假使是这样，这并不一定意味着第二个方程会得到更好的1997年的预测值。到目前为止，我们只能说，根据一直到1996年的数据，通货膨胀的一个滞后有助于解释失业率的变化。

为了得出1997年的预测值，我们需要知道1996年的 $unem$ 和 inf。它们分别是5.4和3.0。因此，从方程（18.48）得出的 $unem_{1997}$ 的预测值是 $1.572+0.732(5.4)$，或者大约5.52，从方程（18.49）得出的预测值是 $1.304+0.647(5.4)+0.184(3.0)$，或者大约5.35。而美国1997年的实际失业率是4.9，所以这两个方程都过高地预测了实际的失业率。而第二个方程的预测值相对来说更好一些。

我们可以很容易地求出95%的预测区间。我们做 $unem_t$ 对 $(unem_{t-1}-5.4)$ 和 $(inf_{t-1}-3.0)$ 的回归，得到截距为5.35——它作为预测值，已经计算过了——且 $\mathrm{se}(\hat{f}_n)=0.137$。所以，由于 $\hat{\sigma}=0.883$，有 $\mathrm{se}(\hat{e}_{n+1})=[(0.137)^2+(0.883)^2]^{1/2}\approx0.894$。从式（18.47）中得出的95%的预测区间是 $5.35\pm1.96(0.894)$，或者说大约是 $[3.6, 7.1]$。这是个很宽的区间，而1997年的实际值4.9正好在这个区间之内。和我们预料的一样，u_{n+1} 的标准误是0.883，$\mathrm{se}(\hat{e}_{n+1})$ 占很大一部分。

一个专业预测人员常常必须对每个时期都作出预测值。比如，在时间 n，她或他要预测 y_{n+1}。接着，当 y_{n+1} 和 z_{n+1} 可以知道后，他或她又要预测 y_{n+2}。即使预测者已经决定采用模型（18.43）了，在预测 y_{n+2}，他还要在两个选择之间作出抉择。第一种选择是用 $\hat{\delta}_0+\hat{\alpha}_1y_{n+1}+\hat{\gamma}_1z_{n+1}$，其中的参数是利用前 n 次观测估计得到的。第二种选择是利用所有的 $n+1$ 次观测重新估计这些参数，然后用相同的公式去预测 y_{n+2}。为了在接下来的时期进行预测，我们一般可以用由前 n 次观测值估计得到的参数，或者，也可以在每次得到新的数据后利用全部观测值更新回归参数。虽然第二种方法需要进行更多的运算，但增加的工作量相对来说是次要的，它可能会（虽然并不一定）更好一些，因为这些回归系数至少在一定程度上根据新的数据进行了调整。

下面我们就举个具体的例子。假使我们希望预测1998年的失业率，采用的模型中只有 $unem$ 和 inf 的一期滞后。那么，第一种方法是，把1997年
598 的失业率和通货膨胀的数值代入式（18.49）的右边；因为在1997年，$unem=4.9$，$inf=2.3$，$unem_{1998}$ 的预测值大约是4.9。（它与1997年的失业率是相同的，这不过是个巧合。）第二种方法是，加进1997年的观测值，重新估计方程，然后再利用新的方程进行预测（见习题18.15）。

方程（18.43）中的模型被称为**向量自回归模型**［vector autoregressive (VAR) model］。我们在第11章学习了什么是自回归模型：建立只有一个序列 $\{y_t\}$ 的模型，这个序列由它自己的过去值组成。在向量自回归模型中，我们对几个序列的过去建立这些序列的模型——这就是“向量”这个词的来

源。如果我们有两个序列 y_t 和 z_t，向量自回归模型就可以由下面的方程组成：

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + \alpha_2 y_{t-2} + \gamma_2 z_{t-2} + \cdots \tag{18.50}$$

和

$$z_t = \eta_0 + \beta_1 y_{t-1} + \rho_1 z_{t-1} + \beta_2 y_{t-2} + \rho_2 z_{t-2} + \cdots$$

其中，在给定关于 y 和 z 的过去信息时，每个方程都有期望值为零的误差。在方程（18.43）中——以及在我们估计过的方程（18.49）的例子中——我们假定了，只要每个变量有一个滞后就可以反映出所有的动态。（$unem_{t-2}$ 和 inf_{t-2}联合显著性的 F 检验，也证实了每个变量只需要有一期滞后。）

就像例18.8所展示的那样，VAR方程在预测方面很有用。在很多情况下，我们对只预测一个变量感兴趣，这时，我们只需要估计和分析有关 y 的方程。我们当然也可以在方程（18.50）中加入其他的滞后变量，比如说 w_{t-1}，w_{t-2}，…。只要方程中包含了所有变量的足够多的滞后，而且方程满足时间序列回归的同方差假定，OLS就可以有效地估计出这些方程。

像式（18.50）这样的方程，使得我们在掌握了过去的 y 之后，可以检验过去的 z 是否有助于预测 y_t。一般来说，我们称 z 是 y 的**葛兰杰原因**（Granger causes），如果

$$\mathrm{E}(y_t | I_{t-1}) \neq \mathrm{E}(y_t | J_{t-1}) \tag{18.51}$$

式中，I_{t-1}包含所有关于 y 和 z 的过去的信息，而 J_{t-1}只包括 y 的过去的信息。当式（18.51）成立时，除了过去的 y 以外，过去的 z 对于预测 y_t 也是有用的。“葛兰杰原因”中的“原因”一词一定要谨慎理解。z 是 y 的“原因”惟一的含义就是式（18.51）。尤其是，它并没有说 y 和 z 之间有同期因果关系，所以，它并不能帮助我们判断在关于 y_t 和 z_t 之间的方程中，z_t 是外生的还是内生的。[**葛兰杰因果关系**（Granger causality）不适用于纯粹的横截面分析也是这个缘故。]

一旦假定了一个线性模型，并决定了在$\mathrm{E}(y_t | y_{t-1}, y_{t-2}, \cdots)$中包含多少个 y 的滞后，就可以很容易地检验 z 不是 y 的葛兰杰原因这一虚拟假设了。具体一点，假设$\mathrm{E}(y_t | y_{t-1}, y_{t-2}, \cdots)$与三个滞后有关：

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + u_t$$
$$\mathrm{E}(u_t | y_{t-1}, y_{t-2}, \cdots) = 0$$

599 现在，在 z 不是y 的葛兰杰原因这一虚拟假设下，我们往方程中添加的 z 的任何滞后都应该有零总体系数。如果添加 z_{t-1}，就只要对 z_{t-1}做一个 t 检验。如果添加 z 的两个滞后，就可以用 F 检验来看看下面方程中的z_{t-1} 和 z_{t-2}的联合显著性：

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + u_t$$

（如果有异方差，我们可以使用异方差一稳健形式的检验。在 H_0 下不会有序列相关存在，因为模型是动态完整的。）

有一个实际的问题：我们如何决定 y 和 z 的哪些滞后要包含在模型中呢？第一步，我们估计 y 的一个自回归模型，并运用 t 和 F 检验来决定应该有多少个 y 的滞后出现在模型中。对于年度数据，滞后的数量一般很小，比如 1 或 2。对于季度或月度数据，滞后的数量往往多很多。一旦 y 的自回归模型确定下来，我们就可以检验 z 的滞后了。z 的滞后数量的选择不怎么重要，因为当 z 不是 y 的葛兰杰原因时，z 的滞后值的集合不可能是显著的。对于年度数据，一般只要 1 个或 2 个滞后就够了；对于季度数据，通常要 4 个或 8 个；对于月度数据，可能要 6 个或 12 个，如果数据足够多也可以是 24 个。

在方程（18.49）中，我们实际上已经完成了检验葛兰杰因果关系的一个例子。拟合得最好的自回归模型是AR(1)模型。我们在方程（18.49）中加进的通货膨胀的一个滞后是很显著的，因此，通货膨胀是失业率的葛兰杰原因。

葛兰杰因果关系有一个推广了的定义，它也是经常用得到的。令 w_t 为第三个序列（或者，它也可以代表几个增加进来的序列）。于是，如果式（18.51）成立，z 就是 y 的以 w 为条件的葛兰杰原因。需要注意的是，此时的 I_{t-1} 包含 y、z 和 w 的过去的信息，而 J_{t-1} 只包括关于 y 和 w 的过去信息。z 是 y 的葛兰杰原因，但却不是 y 的以 w 为条件的葛兰杰原因，这种情况也是有可能的。当虚拟假设是 z 不是 y 的以 w 为条件的葛兰杰原因时，在 y 既依赖于滞后的 z 也依赖于滞后的 w 的一个模型中，检验滞后的 z 的显著性，就可以检验上面的虚拟假设。例如，为了检验货币供给的增长是否是实际 GDP 增长的以利率为条件的变化的葛兰杰原因，我们可以做 $gGDP$ 对 $gGDP$ 的滞后、对 Δint 和 gM 的回归，然后对 gM 的滞后做显著性检验。［参见 Stock 和 Watson（1989）等。］

超前一步预测的比较

在几乎所有的预测问题中，都有几种比较好的预测方法可供选择。即使我们把注意力限定在回归模型上，仍然有较多种可能类型。应该在模型中包括哪些变量，用多少个滞后呢？应该用对数形式还是水平形式，或者是一阶差分呢？

为了在预测方法方面作出决定，我们需要一个规则来确定哪一种是最合适的。宽泛地讲，可以归纳为**样本内准则**（in-sample criteria）和**样本外准则**（out-sample criteria）两种规则。在回归的框架中，样本内准则包括 R-平方和校正的 R-平方。另外还有很多种**模型选择统计量**（model selection statistics），在这里就不介绍了。［参见 Ramanathan（1995，第 4 章）。］

对于预测来说，使用样本外准则更好一些，因为预测本质上是一个样本外问题。一个模型也许在用于估计其参数的样本中对 y 拟合得比较好，但在用于预测时未必好。一个样本外的比较方法大致为，用样本的前一部分去

600 估计模型中的参数，然后用样本中余下的部分来判断它的预测能力。这模拟了我们在不知道变量的将来值时的所要做的事情。

假设有 $n+m$ 次观测，其中的前 n 次观测被我们用来估计模型中的参数，其余的 m 次观测用于预测。令 $\hat{f}_{n+h}$ 代表 $h=0, 1, \cdots, m-1$ 时 y_{n+h+1} 的超前一步预测值。这 m 个预测误差是 $\hat{e}_{n+h+1}=y_{n+h+1}-\hat{f}_{n+h}$（这里，$h=0, 1, \cdots, m-1$——译者注）。当预测超出了样本时，怎样衡量模型预测得有多好呢？有几种常见的衡量标准。第一种是**误差的均方根**（root mean squared error，RMSE）：

$$\mathrm{RMSE}=\left(m^{-1}\sum_{h=0}^{m-1}\hat{e}_{n+h+1}^{2}\right)^{1/2} \tag{18.52}$$

它实质上就是预测误差的样本标准差（未作任何自由度调整）。如果对两种或多种预测方法计算 RMSE，我们会欢迎有最小样本外 RMSE 的方法。

另一种衡量标准是**绝对误差均值**（mean absolute error，MAE），就是预测误差的绝对值的平均：

$$\mathrm{MAE}=m^{-1}\sum_{h=0}^{m-1}\left|\hat{e}_{n+h+1}\right| \tag{18.53}$$

和前一种标准一样，我们更欢迎 MAE 较小的预测方法。其他标准包括最大绝对预测误差的最小化等。

例 18.9　失业率预测的样本外比较

在例 18.8 中，我们发现方程（18.49）比方程（18.48）更好地拟合了我们的样本。而且，至少在预测 1997 年的失业率方面，有滞后的通货膨胀的模型效果更好一些。现在，利用到 1989 年为止的数据，我们来估计这两个模型，剩下的 1990—1996 年的数据用于样本外比较。这里有 7 次样本外观测（准确地说，$n=41$，$m=7$）。在 AR(1)模型中，RMSE＝0.632，MAE＝0.515。在加进了滞后的通货膨胀的模型中，RMSE＝0.550，MAE＝0.362。因此，无论根据哪种模型，包含了 inf_{t-1} 的模型都产生了更好的 90 年代的样本外预测值。在这种情况下，样本内和样本外准则选择了同一个模型。

还有一种方法，与只利用前 n 次观测来估计模型的参数不同的是，我
601 们也可以在每次增加了一次观测后都重新估计模型，然后再用新模型来预测下一个时期的值。

超前多步预测

提前多于一个时期进行预测一般比只提前一个时期要难一些。我们可以正式表述为：假使我们要在时间 t 和更早的时间 s（$s<t$）分别预测 y_{t+1}，那

么，$\mathrm{Var}[y_{t+1}-\mathrm{E}(y_{t+1}|I_t)]\leqslant\mathrm{Var}[y_{t+1}-\mathrm{E}(y_{t+1}|J_t)]$，而且，不等式经常是严格不等的。这里，我们不作一般性的证明，只从直观上考虑一下它的含义：预测 y_{t+1} 的预测误差方差，在利用较少的信息做预测时要来得更大一些。

如果 $\{y_t\}$ 遵循一个AR(1)模型（包括随机游走模型，而且可能有漂移），可以比较容易地证明，随着预测范围的延伸，预测误差的方差将变大。模型为

$$
\begin{aligned}
&y_t=\alpha+\rho y_{t-1}+u_t\\
&\mathrm{E}(u_t|I_{t-1})=0,\ I_{t-1}=\{y_{t-1},y_{t-2},\cdots\}
\end{aligned}
$$

式中，以 I_{t-1} 为条件的 $\{u_t\}$ 有恒定方差 σ^2。在时间 $t+h-1$，预测 y_{t+h} 的值是 $\alpha+\rho y_{t+h-1}$，预测误差是 u_{t+h}。因此，超前一步预测的方差是 σ^2。为了求出超前多步预测值，我们利用反复迭代：

$$
\begin{aligned}
y_{t+h}=&(1+\rho+\cdots+\rho^{h-1})\alpha+\rho^h y_t\\
&+\rho^{h-1}u_{t+1}+\rho^{h-2}u_{t+2}+\cdots+u_{t+h}
\end{aligned}
$$

在时间 t，对于所有的 $j\geqslant 1$，u_{t+j} 的期望值为零。所以

$$\mathrm{E}(y_{t+h}|I_t)=(1+\rho+\cdots+\rho^{h-1})\alpha+\rho^h y_t \tag{18.54}$$

而且，预测误差为 $e_{t,h}=\rho^{h-1}u_{t+1}+\rho^{h-2}u_{t+2}+\cdots+u_{t+h}$，它是一系列不相关随机变量的和，所以和的方差就等于方差的和：$\mathrm{Var}(e_{t,h})=\sigma^2[\rho^{2(h-1)}+\rho^{2(h-2)}+\cdots+\rho^2+1]$。因为 $\rho^2>0$，每个项乘以 σ^2 后都是正的，所以预测误差方差随着 h 的上升而增大。当 $\rho^2<1$ 时，预测方差收敛于 $\sigma^2/(1-\rho^2)$，它正好是 y_t 的非条件方差。在随机游走($\rho=1$)的情况下，$f_{t,h}=\alpha h+y_t$，而且 $\mathrm{Var}(e_{t,h})=\sigma^2 h$，所以，随着预测范围 h 的延伸，预测方差无极限地增大。这说明，预测一个随机游走的很远的将来是很困难的，无论是否带有漂移。比方说，对越远的将来的利率所做预测，准确率越低。

方程（18.54）表明，一旦我们用OLS估计了 ρ，用AR(1)模型进行多步预测就比较容易了。在时间 n，y_{n+h} 的预期值为

$$\hat{f}_{n,h}=(1+\hat{\rho}+\cdots+\hat{\rho}^{h-1})\hat{\alpha}+\hat{\rho}^h y_n \tag{18.55}$$

除非 $h=1$，否则要想计算预测区间会更困难一些，因为很难求出 $\hat{f}_{n,h}$ 的标准误。不过，与误差项的标准差相比，$\hat{f}_{n,h}$ 的标准误通常还算比较小，而前者可以被估计为 $\hat{\sigma}[\hat{\rho}^{2(h-1)}+\hat{\rho}^{2(h-2)}+\cdots+\hat{\rho}^2+1]^{1/2}$，这里的 $\hat{\sigma}$ 是通过AR(1)估计得出的回归的标准误。我们可以用这个结果来求出一个渐近的置信区间。比如，当 $h=2$ 时，一个近似的95%的置信区间（对于较大的 n）为

$$\hat{f}_{n,2}\pm 1.96\hat{\sigma}(1+\hat{\rho}^2)^{1/2} \tag{18.56}$$

602 因为我们低估了 y_{n+h} 的标准方差，这个置信区间偏窄，但可能不会窄很多，尤其是在 n 很大的时候。

一个很正统但很有用的方法是，对每个预测范围估计不同的模型。例如，假使我们希望提前两个时期预测 y。如果 I_t 只取决于一直到时间 t 为止的 y，我们就可以假定 $\mathrm{E}(y_{t+2}|I_t)=\alpha_0+\gamma_1 y_t$［我们在前面见到过，如果 $\{y_t\}$ 遵循一个 AR(1) 模型，它就会成立。］我们可以通过做 y_t 对截距和 y_{t-2} 的回归来估计 α_0 和 γ_1。假使这个方程中的误差包含序列相关——相邻时期的误差是相关的——我们还是可以得到 α_0 和 γ_1 的一致且渐近正态的估计值。那么，在时间 n，y_{n+2} 的预测值是 $\hat{f}_{n,2}=\hat{\alpha}_0+\hat{\gamma}_1 y_n$。另外，很重要的是，这个回归的标准误刚好是我们在计算预测的置信区间时所需要的。不过，为了得到 $\hat{f}_{n,2}$ 的标准误，超前一步预测的方法要求我们求出一个序列相关—稳健标准误，就是我们在 12.5 节介绍过的那种。随着 n 的增大，标准误趋于零，而误差的方差是恒定的。因此，我们可以使用式（18.56）并用通过做 y_t 对 y_{t-2} 回归得到的 SER 替代 $\hat{\sigma}(1+\hat{\rho}^2)^{1/2}$，就可以得出一个渐近的区间。应该注意的是，我们在这里仍然忽略了 $\hat{\alpha}_0$ 和 $\hat{\gamma}_1$ 中的估计误差。

我们也可以用更复杂的自回归模型来计算超前多步的预测值。比如说，假设 $\{y_t\}$ 遵循一个 AR(2) 模型，并打算在时间 n 预测 y_{n+2}。于是，$y_{n+2}=\alpha+\rho_1 y_{n+1}+\rho_2 y_n+u_{n+2}$。所以，

$$\mathrm{E}(y_{n+2}|I_n)=\alpha+\rho_1\mathrm{E}(y_{n+1}|I_n)+\rho_2 y_n$$

上式可以写为

$$f_{n,2}=\alpha+\rho_1 f_{n,1}+\rho_2 y_n$$

这样，一旦得到了超前一步预测值，我们就可以求出在时间 n 的超前两步预测值了。如果 AR(2) 模型的参数已经用 OLS 估计出来了，我们就用下面的式子求出预测值：

$$\hat{f}_{n,2}=\hat{\alpha}+\hat{\rho}_1\hat{f}_{n,1}+\hat{\rho}_2 y_n \tag{18.57}$$

式中，$\hat{f}_{n,1}=\hat{\alpha}+\hat{\rho}_1 y_n+\hat{\rho}_2 y_{n-1}$，我们可以在时间 n 计算出这个预测值。于是把它与 y_n 一道代入式（18.57）求出 $\hat{f}_{n,2}$。对于任何的 $h>2$，求出 AR(2) 模型的超前 h 步预测值很简单，只要反复代换就行了：$\hat{f}_{n,h}=\hat{\alpha}+\hat{\rho}_1\hat{f}_{n,h-1}+\hat{\rho}_2\hat{f}_{n,h-2}$。

对于 VAR 模型，超前多步预测值的计算可以采用类似的思路。下面我们就举个例子说明一下。假使有

$$y_t=\delta_0+\alpha_1 y_{t-1}+\gamma_1 z_{t-1}+u_t \tag{18.58}$$

和

$$z_t=\eta_0+\beta_1 y_{t-1}+\rho_1 z_{t-1}+v_t$$

603 如果想在时间 n 预测 y_{n+1}，可以利用 $\hat{f}_{n,1}=\hat{\delta}_0+\hat{\alpha}_1 y_n+\hat{\gamma}_1 z_n$。同时，$z_{n+1}$ 在时间 n 的预测值（比如说）为 $\hat{g}_{n,1}=\hat{\eta}_0+\hat{\beta}_1 y_n+\hat{\rho}_1 z_n$。现在，假设我们想在时间 n 求出 y 的超前多步预测值。由式（18.58），我们有

$$\mathrm{E}(y_{n+2}|I_n)=\delta_0+\alpha_1\mathrm{E}(y_{n+1}|I_n)+\gamma_1\mathrm{E}(z_{n+1}|I_n)$$

[因为 $E(u_{n+2}|I_n)=0$]，那么，预测值为

$$\hat{f}_{n,2}=\hat{\delta}_0+\hat{\alpha}_1\hat{f}_{n,1}+\hat{\gamma}_1\hat{g}_{n,1} \tag{18.59}$$

这个方程表明，y 的超前两步预测值取决于 y 和 z 的超前一步预测值。一般来讲，我们可以反复使用下面的公式来求出超前多步预测值：

$$\hat{f}_{n,h}=\hat{\delta}_0+\hat{\alpha}_1\hat{f}_{n,h-1}+\hat{\gamma}_1\hat{g}_{n,h-1},\quad h\geqslant 2$$

例 18.10 失业率的超前两年预测值

为了用方程(18.49)提前两年预测失业率——比如说，用截至 1996 年的数据预测 1998 年的失业率——我们需要一个通货膨胀的模型。用滞后 *unem* 和 *inf* 表示 *inf* 的最好模型似乎是简单的 AR(1)模型（模型中的 $unem_{t-1}$ 并不显著）：

$$\begin{aligned}\widehat{inf}_t &= 1.277+0.665\,inf_{t-1}\\ &\quad\ (0.558)\quad\ (0.107)\\ &\quad\ n=48,\ R^2=0.457,\ \bar{R}^2=0.445\end{aligned}$$

如果把 1996 年 *inf* 的值代入到方程中，就会得到 1997 年 *inf* 的预测值：$\widehat{inf}_{1997}=3.27$。接着，我们把它和预测值 $\widehat{unem}_{1997}=5.35$（在前面得到的）一道代入方程（18.59）中来预测 $unem_{1998}$：

$$\widehat{unem}_{1998}=1.304+0.647(5.35)+0.184(3.27)\approx 5.37$$

记住，这个预测只使用了截至 1996 年的信息。把 1997 年 *unem* 和 *inf* 的值代入式（18.48）得到的 $unem_{1998}$ 的超前一步预测值大约为 4.90。你可以在最近的总统经济报告中找到 1998 年的实际失业率。你会发现，超前一步预测值远比超前多步预测值更接近实际值。

样本外误差均方根或平均绝对误差，不但可以帮助我们选择较好的超前一步预测方法，同样，也可以用于超前多步预测方法的选择。

有趋势、季节性和自积过程的预测

现在，开始预测有趋势、有季节性或有单位根的序列。在第 10 章和第 11 章介绍过，处理回归模型中有趋势应变量或自变量的方法是，在模型中加进时间趋势，其中最常见的是线性趋势。趋势也可以加在预测方程中，但一定要慎重行事。 604

在最简单的情况下，假使 $\{y_t\}$ 有一个线性趋势，但它在趋势周围的变化是不可预测的。那么，可以写出

$$y_t=\alpha+\beta t+u_t,\ \mathrm{E}(u_t|I_{t-1})=0,\ t=1,2,\cdots \tag{18.60}$$

式中，I_{t-1}包含到时间 $t-1$ 为止所有观测到的信息（它至少包含过去的 y 的信息）。对于任何 $h \geqslant 1$，我们如何在时间 n 预测 y_{n+h}呢？这很简单，因为 $\mathrm{E}(y_{n+h}|I_n)=\alpha+\beta(n+h)$。预测误差方差是$\sigma^2=\mathrm{Var}(u_t)$（假定方差在不同时间是相同的）。利用前 n 次观测，我们用 OLS 估计出 α 和β，那么在时间 n 我们对 y_{n+h}的预测为$\hat{f}_{n,h}=\hat{\alpha}+\hat{\beta}(n+h)$。换句话说，只要把与 y 对应的时期值代入估计趋势函数中。例如，如果用 BARIUM.RAW 中的 $n=131$ 次观测来预测每月美国从中国进口氯化钡的进口量，得到 $\hat{\alpha}=249.56$ 和 $\hat{\beta}=5.15$。样本期间到 1988 年 12 月为止，所以 6 个月后从中国的进口量的预测值为249.56+5.15(137)=955.11（以短吨计）。实际上，1988 年 12 月的值是 1 087.81，它比预测到的 6 个月后的值要大一些。图 18.2 展示了这个序列及估计到的趋势线。

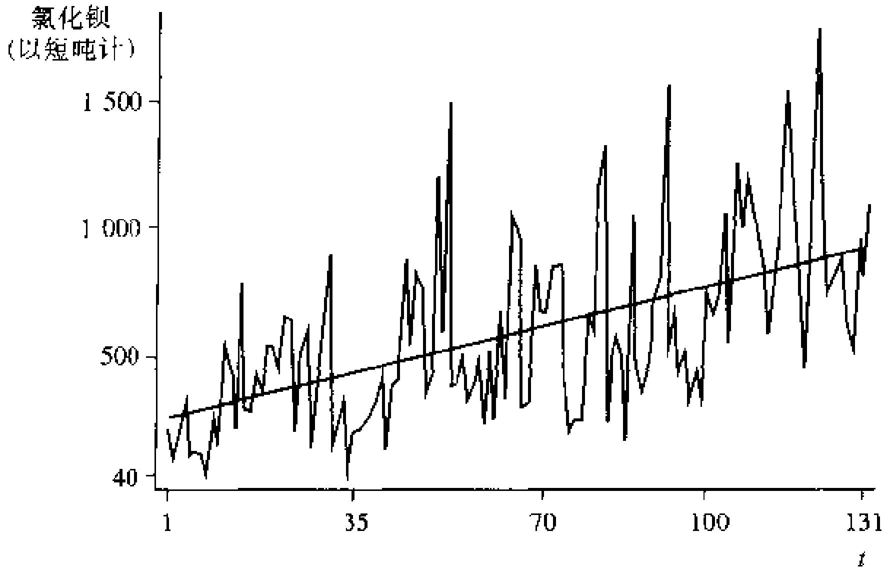


图 18.2　来自中国的美国氯化钡的进口量和估计到的它的线性趋势线，249.56+5.15t

605 我们在第 10 章谈到，大多数经济的时间序列有恒定的增长率，我们至少可以渐近地这样认为。这样，$\log(y_t)$就有一个线性的时间趋势。假使我们用 n 次观测求出下面的方程：

$$\widehat{\log}(y_t)=\hat{\alpha}+\hat{\beta}t,\quad t=1,2,\cdots,n \tag{18.61}$$

那么，为了预测$\log(y)$在将来的任何时间 $n+h$ 的值，只要将 $n+h$ 代入趋势方程中。但这并不能让我们预测出 y，而在很多时候我们偏偏又想知道它。有些人想通过指数化$\hat{\alpha}+\hat{\beta}(n+h)$来计算 y_{n+h}的预测值，但这么做是不甚正确的，在 6.4 节我们提出了相应的理由。我们必须处理好隐含在式(18.61)中的误差。最简单的方法是用这 n 次观测做 y_t 对$\exp(\widehat{\log}y_t)$没有截距的回归。令 $\hat{\gamma}$ 为$\exp(\widehat{\log}y_t)$的斜率系数，那么在时间 $n+h$ 的 y 的预测值就是

$$\hat{f}_{n,h}=\hat{\gamma}\exp[\hat{\alpha}+\hat{\beta}(n+h)] \tag{18.62}$$

下面举个例子。如果我们用 NYSE.RAW 中纽约股票交易所指数的最初 687 周的数据，求出 $\hat{\alpha}=3.782$ 和 $\hat{\beta}=0.001\,9$［通过做 log（$price_t$）对一个线性时间趋势的回归］；以上结果表明，指数平均每周上升约 0.2%。当我们做 *price* 对指数化的拟合值的回归时，得到 $\hat{\gamma}=1.018$。现在利用式（18.62）预测 4 周后的价格，也就是样本中的最后一周的价格：$1.018\exp[3.782+0.001\,9(691)]\approx166.12$。实际值是 164.25，所以预测的结果偏高。但这个结果比估计最初 687 周的一个线性时间趋势得到的结果要好得多：第 691 周的预测值是 152.23，明显严重过低预测了。

问题 18.5 假使你为 $\{y_t: t=1, 2, \cdots, 46\}$ 构造了一个线性时间趋势模型，使用的是 1950—1995 年的年度数据。把变量 $year_t$ 定义为从 50（$t=1$ 时）到 95（$t=46$ 时）。如果你估计方程 $\hat{y}_t=\hat{\gamma}+\hat{\delta}year_t$，请把得到的 $\hat{\gamma}$ 和 $\hat{\delta}$ 与 $\hat{y}_t=\hat{\alpha}+\hat{\beta}t$ 中的 $\hat{\alpha}$ 和 $\hat{\beta}$ 进行比较，它们之间有什么不同？从这两个方程中得出的预测值又有什么区别？

虽然趋势模型在预测中很有用处，但我们要慎重使用这种方法，特别是在对有漂移的自积序列预测很远的将来值的时候。考虑下面的一个有漂移的随机游走，就可以发现可能存在的问题了。在时间 $t+h$，我们可以把 y_{t+h} 写为

$$y_{t+h}=\beta h+y_t+u_{t+1}+\cdots+u_{t+h}$$

式中，β 为漂移项（通常 $\beta>0$）；给定 I_t 及恒定的方差 σ^2，每个 u_{t+j} 都有零均值。我们前面已经讲过，y_{t+h} 在时间 t 的预测值为 $\mathrm{E}(y_{t+h}|I_t)=\beta h+y_t$，预测误差方差是 σ^2h。如果我们改用线性趋势模型，情况会如何呢？令 y_0 为这个过程在时间 0 的初始值，我们把它当做非随机的。于是，我们可以写成

$$\begin{aligned}y_{t+h}&=y_0+\beta(t+h)+u_1+u_2+\cdots+u_{t+h}\\&=y_0+\beta(t+h)+v_{t+h}\end{aligned}$$

606 这看起来有些像截距 $\alpha=y_0$ 的线性趋势模型。当误差 v_{t+h} 有零均值时，它的方差为 $\sigma^2(t+h)$。因此，如果在时间 t 用线性趋势 $y_0+\beta(t+h)$ 来预测 y_{t+h}，预测误差方差为 $\sigma^2(t+h)$，而当我们用 $\beta h+y_t$ 时预测误差方差为 σ^2h。预测方差的比为 $(t+h)/h$，如果 t 很大，这个比就会很大。那么，结论显然应该是，我们不应该使用线性时间趋势去预测有漂移的随机游走。（在习题 18.17 中，你将比较通过两种不同的方法得出的美国总生育率的预测值有何不同，一种方法是利用一个时间的三次趋势线，另一种方法利用简单的随机游走模型。）

如果使用过时的数据来估计趋势参数，而且这个过程的趋势线在其后有所移动，那么确定性的趋势会产生比较差的预测结果。有时，外生的冲击——比如说 20 世纪 70 年代的石油危机——可以改变趋势变量的轨迹。如果用一个过时的趋势线预测很远的将来，预测值可能会偏离实际值较大。这个问题不难解决，只需要使用最新的数据来求趋势线的参数。

我们完全可以把趋势和其他模型结合起来进行预测。比如，我们可以在AR(1)模型中加进一个线性趋势，这种方法对于预测有线性趋势而且围绕趋势的是平稳的AR的模型会很有效。

预测有确定的季节性（月度或季度）的过程比较简单。比如，BARIUM.RAW包含美国1978—1988年汽油的每月产量数据。这个序列没有明显的趋势，但它有很强的季节性特点。（汽油的产量在夏天的几个月比在12月高。）在最简单的模型中，我们做 *gas*（以加仑计）对表示11个月份的虚拟变量的回归，这些虚拟变量可以是 *February* 到 *December*。于是，将来的任何一个月的预测值就是截距加上那个月的对应虚拟变量的系数。（对于 *January*，预测值就是回归中的截距。）我们也可以在模型中加进变量的滞后和时间趋势，使它广泛适用于有季节性的一般序列。

对有单位根的过程的预测应该受到足够的重视。在前面，我们求出了以到时间 n 为止的信息为条件的一个随机游走的期望值。在时间 n，为了预测可能有漂移 α 的随机游走的以后的第 h 个时期的值，我们用 $\hat{f}_{n,h}=\hat{\alpha}h+y_n$，其中的 $\hat{\alpha}$ 是 Δy_t 的直到 $t=n$ 时为止的样本均值。（如果没有漂移，我们设 $\hat{\alpha}=0$。）这种方法在模型中增加了单位根。另一种方法是，估计 $\{y_t\}$ 的AR(1)模型，然后运用预测式（18.55）。这种方法没有在方程中加设单位根，但是如果出现了单位根，随着 n 增大，ρ 在概率上收敛于1。不过，ρ 有可能非常不同于1，特别是在样本容量不是很大的时候。哪一种方法能带来更好的样本外预测，是个经验性问题。如果在AR(1)模型中，ρ 小于1，哪怕只小一点，AR(1)模型可能就会提供更好的长期预测值。

一般而言，有两种方法可以用来预测I(1)过程。第一种方法是在模型中加设单位根。对于一次超前一步预测，给定直至时间 t 的信息，我们求出一个用来预测 y 的变化 Δy_{t+1} 的模型。于是，因为 $y_{t+1}=\Delta y_{t+1}+y_t$，故有 $\mathrm{E}(y_{t+1}|I_t)=\mathrm{E}(\Delta y_{t+1}|I_t)+y_t$。因此，$y_{n+1}$ 在时间 n 的预测值为

$$\hat{f}_n=\hat{g}_n+y_n$$

式中，$\hat{g}_n$ 为 Δy_{t+1} 在时间 n 的预测值。典型的做法是用AR模型（它一定是平稳的）或者向量自回归来表示 Δy_t。

607 通过把 y_{n+h} 写成下面的形式，可以把以上方法扩展到超前多步预测中：

$$y_{n+h}=(y_{n+h}-y_{n+h-1})+(y_{n+h-1}-y_{n+h-2})+\cdots+(y_{n+1}-y_n)+y_n$$

或

$$y_{n+h}=\Delta y_{n+h}+\Delta y_{n+h-1}+\cdots+\Delta y_{n+1}+y_n$$

所以，y_{n+h} 在时间 n 的预测值为

$$\hat{f}_{n,h}=\hat{g}_{n,h}+\hat{g}_{n,h-1}+\cdots+\hat{g}_{n,1}+y_n \tag{18.63}$$

式中，$\hat{g}_{n,j}$ 为 Δy_{n+j} 在时间 n 的预测值。比如，我们可以用一个平稳的AR(1)模型表示 Δy_t，然后用式（18.55）求出超前多步预测值（$\hat{\alpha}$ 和 $\hat{\rho}$ 是通过做 Δy_t 对 Δy_{t-1} 的回归得到的，而且 y_n 由 Δy_n 代替），然后把它们代入式（18.63）。

第二种预测I(1)变量的方法是，利用$\{y_t\}$的一般的AR或VAR模型。这次没有加设单位根。下面举个例子。如果用AR(2)模型

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t \tag{18.64}$$

那么，$\rho_1 + \rho_2 = 1$。如果我们把$\rho_1 = 1 - \rho_2$代进去并整理一下，得到$\Delta y_t = \alpha - \rho_2 \Delta y_{t-1} + u_t$，这是个差分形式的平稳的AR(1)模型，它使得我们又回到了前面介绍的那种方法。可以直接用OLS估计式（18.64）。这种回归的一个优点是，可以使用ρ_2的通常的t统计量来判断y_{t-2}是否是显著的。（这要求同方差假定成立；如果它不成立，可以使用异方差—稳健统计量。）我们在这里不打算证明这一点，但从直观来看，把方程改写为$y_t = \alpha + \gamma y_{t-1} - \rho_2 \Delta y_{t-1} + u_t$（$\gamma = \rho_1 + \rho_2$），就可以看出来了。即令$\gamma = 1$，$\rho_2$是一个平稳的、弱相依的序列$\{\Delta y_{t-1}\}$的系数的相反数。因为这个回归的结果与式（18.64）中的相同，我们可以直接应用它。

作为一个具体的例子，我们来估计一下FERTIL3.RAW中的总生育率的一个AR（2）模型，使用截至1979年的观测。（在习题18.17中，要求你用这个模型来进行预测，这也是我们把样本的后一部分观测留下来的原因。）

$$\begin{aligned}\hat{gfr}_t = {} & 3.22 + 1.272\,gfr_{t-1} - 0.311\,gfr_{t-2} \\ & (2.92) \quad (0.120) \qquad (0.121) \\ & n = 65,\ R^2 = 0.949,\ \bar{R}^2 = 0.947\end{aligned} \tag{18.65}$$

第二个滞后的t统计量大约是-2.57，它在1%的水平上是统计上异于1的。（第一个滞后也有一个非常显著的t统计量，它也有一个渐近的t分布，道理与ρ_2的一样。）R-平方，无论是经过调整的还是没有经过调整的，作为拟合优度指标都不能提供特别有价值的信息，因为gfr明显包含了单位根，所以，讨论gfr的方差中有多少被我们解释了基本上没有什么意义。

608 式（18.65）中的两个滞后的系数之和为0.961，它很接近于1，统计上也不是异于1的（将扩展的迪基-富勒检验用于方程$\Delta gfr_t = \alpha + \theta gfr_{t-1} + \delta \Delta gfr_{t-1} + u_t$就可以验证）。即使我们没有施加单位根条件，仍然可以像在前面介绍过的那样使用式（18.65）来进行预测。

在结束这一部分之前，对于有I(1)变量的向量自回归模型情况下的预测，我们提出一种有潜力的改进方法。假设$\{y_t\}$和$\{z_t\}$都是I(1)过程。取得y的预测值的一种方法是，估计变量Δy_t和Δz_t的一个双变量自回归模型，然后用式（18.63）算出超前一步或多步预测值；这种方法在本质上与前面介绍的第一种方法相同。但是，如果y_t和z_t是协积的，在用于预测Δy的信息集里就增加了更多平稳的、稳定的变量：$y_t - \beta z_t$的滞后，式中的β为协积系数。一个简单的误差纠正模型为

$$\begin{aligned}&\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_1 \Delta z_{t-1} + \delta_1 (y_{t-1} - \beta z_{t-1}) + e_t \\ &\mathrm{E}(e_t \mid I_{t-1}) = 0\end{aligned} \tag{18.66}$$

为了预测y_{n+1}，我们用一直到n的观测来估计协积系数β，然后用OLS估计误差纠正模型的参数，方法见于18.4节。那么，预测Δy_{n+1}就很简单了：

只要把 Δy_n，Δz_n 和 $y_n - \hat{\beta}z_n$ 代入上面的方程。在求出 Δy_{n-1} 后，再把它和 y_n 加起来就行了。

重新整理误差纠正模型，我们可以写成

$$y_t = \alpha_0 + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \tag{18.67}$$

它是 y_t 和 z_t 的 VAR 模型中的第一个方程，其中 $\rho_1 = 1 + \alpha_1 + \delta$，$\rho_2 = -\alpha_1$ 等等。请注意，上面的方程中有 5 个参数，和误差纠正模型中的一样多。尤为关键的是，就预测的目的而言，水平形式的 VAR 模型和误差纠正模型在本质上是一样的。但是，在更一般的误差纠正模型中就没有这样的结论了。比如，若在式(18.66)中 $\alpha_1 = \gamma_1 = 0$，但我们还有一个误差纠正项 $\delta_2(y_{t-2} - \beta z_{t-2})$。于是，这个误差纠正模型只涉及 4 个参数，而式（18.67）——y 和 z 的滞后的阶数是一样的——包含了 5 个参数。所以，误差纠正模型能节约参数的个数，它一般比水平形式的 VAR 模型使用更少的参数。

如果 y_t 和 z_t 是I(1)但又不是协积的，那么适当的模型将是没有误差纠正项的式（18.66）。它可被用来预测 Δy_{n+1}，我们把它与 y_n 加起来就可以用于预测 y_{n+1} 了。

小结

这一章所介绍的时间序列的专题程序化地广泛应用于实证宏观经济学、实证金融和其他很多实际应用领域。我们是从说明如何解释和估计无限分布滞后模型开始的。这些模型提供了灵活的滞后分布，而且比类似的有限分布滞后模型使用了更少的参数。几何分布滞后模型，或者更为一般的有理分布滞后模型，是最常见的无限分布滞后模型。用于简单动态方程的标准计量经济学程序，就可以用来估计它们。

609 在时间序列计量经济学中，单位根的检验已经是极为常见的了。如果一个序列有单位根，那么在很多情况下，通常的大样本正态渐近性不再生效。此外，一个单位根过程有这样的一种性质，即它的**新生值**（innovation）有长久的影响，这个问题本身就很值得研究。虽然有很多种单位根检验的方法，迪基-富勒检验——以及它的扩展，即扩展的迪基-富勒检验——大概是最常见的，也是最容易操作的。通过在迪基－富勒回归中添加一个趋势变量，我们就可以在做单位根检验的同时解释线性趋势了。

当一个I(1)序列 y_t 对另一个I(1)序列 x_t 回归时，即使这些序列不包含明显的趋势，我们仍非常担心谬误回归的问题。这一点已经在随机游走的情况下研究得比较彻底了：即使两个随机游走是相互独立的，以通常的临界值为基础的、斜率系数的显著性 t 检验，检验被拒绝的机会比名义上的大小还多。此外，R^2 趋近于一个随机变量，而不是零（如果我们做 y_t 的差分对 x_t 的差分的回归，它就会趋近于零）。

关于I(1)变量的回归不是谬误的一种重要情况是，这些序列是协积的。这意味着这两个I(1)变量的线性函数是I(0)。如果 y_t 和 x_t 是I(1)，但 y_t x_t 是I(0)，y_t 和 x_t 之间就不会分开任意远。如果虚拟假设是没有协积，对立假设是有协积，有一些简单的检验办法，其中一种是，对从静态回归中得到的残差运用迪基-富勒单位根检验。也有一些简单的协积参数的估计量，它们有是近似标准正态分布的 t 统计量（也有渐近生效的置信区间）。我们在18.4节介绍了超前和滞后估计量。

y_t 和 x_t 之间的协积意味着，在一个关于 Δy_t 和 Δx_t 的模型中会有一些误差纠正项；这些误差纠正项是 $y_t-\beta x_t$ 的滞后，其中的 β 是协积系数。一种简单的两步估计法可以用来估计误差纠正模型。首先，用一个静态回归（或者超前和滞后回归）来估计 β。然后，用OLS来估计一个一阶差分形式的简单动态模型，模型中应包含误差纠正项。

18.5节介绍了预测方面的问题，侧重于以回归为基础的预测方法。静态模型，或更一般地说，解释变量与应变量都是同期的模型，不在我们的考虑之列，否则，还需要预测解释变量。如果我们把未知的、假设的解释变量的将来值代进去，就得到了一个条件预测值。非条件预测，无异于构造一个 y_t 的模型，其中把 y_t 表示成预测时已观测到的过去信息的函数。动态回归模型，包括自回归和向量自回归模型，可以直接用计算机程序来计算。除了超前一步点预测以外，我们还讨论了预测区间的构造问题，它与预报区间很相似。

有很多种选择预测方法的准则，最为常见的效果指标是误差的均方根和平均绝对误差。这两者都是对平均预测误差的大小的一种评估。用样本外预测值来计算这些指标更好些，能够提供更有价值的信息。

超前多步预测给我们带来了新的挑战，容易出现大的预测误差方差。尽管如此，对于自回归和向量自回归这样的模型，超前多步预测值是可以计算

610 的，我们也可以求出近似的预测区间。

预测有趋势的序列和I(1)序列过程中需要特别小心。对有确定趋势的过程，可以在回归模型中加入时间趋势，模型中还可以有变量的滞后，然后加以预测。确定的趋势有一个不足，它的长期预测的效果可能不太好，这是因为，一旦估计出模型，线性趋势表示不断增加或减少。预测I(1)过程的一种典型的方法是，预测这个过程的差分，然后把变量的水平值加到预测到的差分上去。另外，可以采用序列水平形式的向量自回归模型。如果序列之间是协积的，可以用误差纠正模型。

关键术语

增广迪基-富勒检验
协积
前导与滞后估计量
损失函数

条件预测（值）
迪基-富勒分布
迪基-富勒（DF）检验
恩格尔-葛兰杰两步法
误差纠正模型
指数平滑
预测误差
预测区间
几何（或考依克）分布滞后
葛兰杰因果关系
样本内准则
无限分布滞后（IDL）模型
信息集
鞅
鞅差分序列
平均绝对误差（MAE）
超前多步预测
超前一步预测
样本外准则
点预测
有理分布滞后（RDL）模型
误差的均方根（RMSE）
谬误回归问题
非条件预测
单位根
向量自回归（VAR）模型

习　题

18.1　考虑 $k=2$ 时的方程（18.15）。用工具变量法估计 γ_h 和 ρ，你将用什么作为 y_{t-1} 的工具变量？

18.2　一个有趣的、把 y_t 和 x_t 的期望值（x_t^*）联系起来的经济模型推出了。在这个计量经济学模型中，x_t 的期望值是以在时间 $t-1$ 所观测到的所有信息为条件的：

$$y_t=\alpha_0+\alpha_1x_t^*+u_t \tag{18.68}$$

它是一个滞后的应变量的计量经济学模型。对 $\{u_t\}$ 的一个很自然的假定是 $\mathrm{E}(u_t|I_{t-1})=0$，式中的 I_{t-1} 代表在时间 $t-1$ 的关于 y 和 x 的所有信息；这意味着 $\mathrm{E}(y_t|I_{t-1})=\alpha_0+\alpha_1x_t^*$。为了完成这个模型，需要一个关于期望 x_t^* 是如何形成的假定。我们在 11.2 节讲过一个适应性预期的简单例子，在那里有 $x_t^*=x_{t-1}$。一个更复杂一些的适应性预期机制为

$$x_t^*-x_{t-1}^*=\lambda(x_{t-1}-x_{t-1}^*) \tag{18.69}$$

611 式中，$0<\lambda<1$。这个方程的含义是，期望的变化要根据上一期的实现值是高于还是低于相应的预期值来作出反应。假定 $0<\lambda<1$，说明期望的变化是上一期误差的一个比例。

（i）证明以上两个方程意味着

$$y_t=\lambda\alpha_0+(1-\lambda)y_{t-1}+\lambda\alpha_1x_{t-1}+u_t-(1-\lambda)u_{t-1}$$

[提示：把方程（18.68）滞后一个时期，再乘以（$1-\lambda$），把得到的方程从方程（18.68）中减掉。然后再利用方程（18.69）。]

（ii）在 $\mathrm{E}(u_t|I_{t-1})=0$ 下，$\{u_t\}$ 是序列不相关的。对误差 $v_t=u_t-(1-$

$\lambda)u_{t-1}$米讲，这意味着什么？

（iii）如果把（i）部分中的方程改写为

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + v_t$$

我们怎样才能一致地估计 β_j 呢？

（iv）给定 β_j 的一致估计值，你怎样去一致地估计 λ 和 α_1 呢？

18.3　假使$\{y_t\}$和$\{x_t\}$都是I(1)序列，但对于某个不等于零的 β，$y_t - \beta x_t$ 是I(0)。证明对于任何的 $\delta \neq \beta$，$y_t - \delta x_t$ 一定是I(1)。

18.4　考虑方程（18.37）中的误差纠正模型。证明如果你加进另一个误差纠正项 $y_{t-2} - \beta x_{t-2}$，这个方程就会有完全多重共线性。[提示：证明 $y_{t-2} - \beta x_{t-2}$ 是 $y_{t-1} - \beta x_{t-1}$，Δx_{t-1} 和 Δy_{t-1} 的一个完全线性函数。]

18.5　假使过程$\{(x_t, y_t): t = 0, 1, 2, \cdots\}$满足方程

$$y_t = \beta x_t + u_t$$

和

$$\Delta x_t = \gamma \Delta x_{t-1} + v_t$$

式中，$\mathrm{E}(u_t | I_{t-1}) = \mathrm{E}(v_t | I_{t-1}) = 0$；$I_{t-1}$包含所有在时间 $t-1$ 及以前的关于 x 和 y 的信息；$\beta \neq 0$，且$|\gamma| < 1$[使得 x_t 是I(1)，并因此使得 y_t 也是I(1)]。证明：从这两个方程可以推出以下形式的一个误差纠正模型：

$$\Delta y_t = \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + e_t$$

式中，$\gamma_1 = \beta\gamma$；$\delta = -1$；$e_t = u_t + \beta v_t$。[提示：首先从第一个方程的两边减去 y_{t-1}；然后在右边加上并减去一个 βx_{t-1}，并整理；最后，用第二个方程求出包含 Δx_{t-1}的误差纠正模型。]

18.6　利用 VOLAT.RAW 中的月度数据，估计下面的模型：

$$\begin{aligned}\widehat{pcip} = {} & 1.54 + 0.344\,pcip_{-1} + 0.074\,pcip_{-2} \\ & (0.56)\ (0.042)\qquad (0.045) \\ & + 0.073\,pcip_{-3} + 0.031\,pcsp_{-1} \\ & (0.042)\qquad (.013) \\ & n = 544,\ R^2 = 0.174,\ \bar{R}^2 = 0.168\end{aligned}$$

612 式中，$pcip$ 为每月工业产值的百分比变化，它被换算成了年率；$pcsp$ 为标准-普尔 500 指数的百分比变化，也是年率。

(i)如果过去 3 个月的 $pcip$ 是零，且 $pcsp_{-1} = 0$，那么，这个月的工业产值增长的预告值是多少？它是统计上显著的吗？

(ii)如果过去 3 个月的 $pcip$ 是零，但 $pcsp_{-1} = 10$，那么，预告的工业产值的增长又会是多少？

(iii)你认为股票市场对实体经济有什么影响？

18.7　令 gM_t 表示货币供给的年增长，$unem_t$ 表示失业率。假定 $unem_t$ 遵循稳定的AR(1)模型，请详细说明你将如何检验 gM 是否是 $unem$ 的葛兰杰原因。

18.8 假使 y_t 遵循下列模型：

$$y_t = \alpha + \delta_1 z_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

$$\mathrm{E}(e_t | I_{t-1}) = 0$$

式中，I_{t-1}包含在时间 $t-1$ 及以前的 y 和 z 的所有信息。

(i) 证明$\mathrm{E}(y_{t+1} | I_t) = (1-\rho)\alpha + \rho y_t + \delta_1 z_t - \rho\delta_1 z_{t-1}$。[提示：写出 $u_{t-1} = y_{t-1} - \alpha - \delta_1 z_{t-2}$，然后把它代入第二个方程；再把得到的结果代入第一个方程，再取条件均值。]

(ii) 假使你用 n 次观测来估计 α，ρ 和 δ_1，写出用来预测 y_{n+1}的方程。

(iii) 为什么说有 z 的一个滞后和AR(1)序列相关的模型是下面的模型的一个特例？

$$y_t = \alpha_0 + \rho y_{t-1} + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + e_t$$

(iv) 对于利用有AR(1)序列相关的模型进行预测，(iii) 部分有何启示？

18.9 令$\{y_t\}$代表一个I(1)序列。假使 $\hat{g}_n$ 是 Δy_{n+1}的超前一步预测值，令 $\hat{f}_n = \hat{g}_n + y_n$ 为 y_{n+1}的超前一步预测值。说明为什么预测 Δy_{n+1}和 y_{n+1}有相同的预测误差。

计算机习题

18.10 本题利用 WAGEPRC.RAW 中的数据。习题 11.5 给出了 *gprice* 对*gwage* 的一个有限分布滞后模型的估计值，那里用了 *gwage* 的 12 个滞后。

(i) 估计一个 *gprice* 对*gwage* 的简单几何 DL 模型。特别是，要用 OLS 估计方程（18.11）。所估计的即期倾向和长期倾向（LRP）是多少？试勾画所估计的滞后分布。

(ii) 把所估计的即期倾向和 LRP 与习题 11.5 中得到的结果进行比较，并比较一下所估计的滞后分布有何不同。

618 (iii) 现在来估计方程（18.16）中的有理分布滞后模型。描述估计到的滞后分布，并比较这里估计的 IP 和 LRP 与（ii）部分中得到的结果有何不同。

18.11 本题利用 HSEINV.RAW 中的数据。

(i) 检验log(*invpc*)是否有单位根，模型中含有一个线性时间趋势和Δlog(*incpc*$_t$)的两个滞后，显著性水平取为5%。

(ii) 用（i）部分中的方法检验log(*price*)中的单位根。

(iii) 给定（i）和（ii）中的结果，那么检验log(*invpc*)和 log(*price*)之间的协积还有意义吗？

18.12 本题利用 VOLAT.RAW 中的数据。

(i) 估计 *pcip* 的AR(3)模型。然后，再加入第四个滞后，并证明它是非常不显著的。

(ii) 在（i）部分的AR(3)模型中，加进 *pcsp* 的三个滞后来检验 *pcsp* 是

否是 *pcip* 的葛兰杰原因。你有何结论？

(iii) 在 (ii) 部分的模型中，加进 3 个月国债利率 $i3$ 的三个滞后变化。*pcsp* 是 *pcip* 的以过去的 $\Delta i3$ 为条件的葛兰杰原因吗？

18.13　在检验例 18.5 中 *gfr* 和 *pe* 之间的协积的过程中，在方程 (18.32) 中添加 t^2，求出 OLS 残差，并在扩展的 DF 中加进一个滞后。新的结论是什么？这个检验的 5% 的临界值是 -4.15。

18.14　本题利用 INTQRT.RAW 中的数据。

(i) 估计下面的方程：

$$hy6_t = \alpha + \beta hy3_{t-1} + \phi_0 \Delta hy3_t + \phi_1 \Delta hy3_{t-1} + \rho_1 \Delta hy3_{t-2} + e_t$$

并用方程形式报告结果。相对于双侧对立假设，检验 H_0：$\beta = 1$。假定方程中已经有了足够多的超前和滞后，使得 $\{hy3_{t-1}\}$ 在方程中是严格外生的，我们不必担心序列相关。

(ii) 在方程 (18.39) 中的误差纠正模型中，加进 $\Delta hy3_{t-2}$ 和 $(hy6_{t-2} - hy3_{t-3})$。这两项是联合显著的吗？怎样才是适当的误差纠正模型，你有何见解？

18.15　本题利用 PHILLIPS.RAW 中的数据，并加上 1997 年的数据：*unem* 为 4.9，*inf* 为 2.3。

(i) 用截至 1997 年的数据估计式 (18.48) 和式 (18.49) 中的模型。所估计的参数与式 (18.48) 和式 (18.49) 中的结果相比有很大不同吗？

(ii) 用新方程预测 $unem_{1998}$，小数点后保留两位数。从总统经济报告 (1999 年或更新的) 找出 $unem_{1998}$。哪个方程预测得更好？

(iii) 我们在课文中讨论过，用式 (18.49) 预测到的 $unem_{1998}$ 为 4.90。把它与利用截至 1997 年的数据作出的预测相比较。多用一年的数据求得的参数，使我们预测得更好了吗？

(iv) 用式 (18.48) 中的估计模型求出 *unem* 的超前两步预测值。也就是，利用 $\hat{\alpha} = 1.572$，$\hat{\rho} = 0.732$，$h = 2$ 时的方程 (18.55)。与把 $unem_{1997} = 4.9$ 代入式 (18.48) 所求出的超前一步预测值相比，哪一个更好一些？ 61k

18.16　本题利用 BARIUM.RAW 中的数据。

(i) 用前 119 次观测 (也就是不包括最后的 1988 年的 12 个月的观测)，估计线性趋势模型 $chnimp_t = \alpha + \beta t + u_t$。这个回归的标准误是什么？

(ii) 用除了最后 12 个月以外的数据，估计 *chnimp* 的 AR(1) 模型。把这个回归的标准误与 (i) 部分中的标准误相比较，哪一个模型提供了更好的样本内拟合？

(iii) 用 (i) 和 (ii) 部分中的模型计算 1988 年的 12 个月的超前一步预测误差。(每个方法都有 12 个预测误差。) 计算并比较这两种方法的 RMSE 和 MAE。就样本外的超前一步预测而言，哪种方法效果更好？

(iv) 在 (i) 部分的回归中加进月度虚拟变量。它们是联合显著的吗？(当我们检验联合显著性时不必担心误差中的轻度序列相关。)

18.17　本题利用 FERTIL3.RAW 中的数据。

(i) 用横坐标代表时间，画出 gfr 的曲线。在整个样本期间，它包含了明显的向上或向下的趋势吗？

(ii) 利用截至 1979 年的数据，估计 gfr 的立方时间趋势模型（就是做 gfr 对 t，t^2 和 t^3 及截距的回归）。评论一下这个回归的 R 平方。

(iii) 用（ii）部分中的模型，计算 1980—1984 年的超前一步预测误差的 MAE。

(iv) 利用截至 1979 年的数据，做 Δgfr_t 对一个常数的回归。这个常数统计上异于零吗？如果我们假定 gfr_t 遵循一个随机游走，同时也假定漂移项是零，这样做合理吗？

(v) 用随机游走模型预测 1980—1984 年的 gfr：gfr_{n+1} 的预测值就是 gfr_n。求出 MAE。它与（iii）部分中得出的 MAE 有什么差别？你更喜欢哪一种方法？

(vi) 用截至 1979 年的数据，估计 gfr 的AR(2)模型。第二个滞后是显著的吗？

(vii) 用AR(2)模型求出 1980—1984 年的 MAE。这个更一般的模型比随机游走模型样本外预测的效果更好吗？

18.18 本题利用 CONSUMP.RAW 中的数据。

(i) 令 y_t 代表实际个人可支配收入。用截至 1989 年的数据估计下面的模型：

$$y_t = \alpha + \beta t + \rho y_{t-1} + u_t$$

615 并用通常的格式报告结果。

(ii) 用（i）部分估计到的方程预测 1990 年的 y。预测误差是多少？

(iii) 用（i）部分估计到的参数，计算 20 世纪 90 年代的超前一步预测值的 MAE。

(iv) 把 y_{t-1} 从方程中去掉后，计算相同时期内的 MAE。在模型中包含 y_{t-1} 更好些吗？

18.19 本题利用 INTQRT.RAW 中的数据。

(i) 利用除了最后 4 年（16 个季度）以外的所有数据，估计 $\Delta r6_t$ 的AR(1)模型。（我们用差分形式，因为 $r6_t$ 看起来好像有单位根。）用最后 16 个季度的数据，求出 $\Delta r6$ 的超前一步预测的 RMSE。

(ii) 在（i）部分的方程中加入误差纠正项 $spr_{t-1} = r6_{t-1} - r3_{t-1}$。（这相当于假定了协积参数为 1。）计算最后 16 个季度的 RMSE。在这里，误差纠正项对样本外预测有什么帮助吗？

(iii) 现在请你估计协积参数，而不是把它设为 1。再利用最后 16 个季度的数据求出样本外 RMSE。它与（i）和（ii）部分中的结果有什么不同？

(iv) 如果你想要预报的是 $r6$ 而不是 $\Delta r6$，你的结论会有变化吗？请解释为什么。

第 19 章　一个经验项目的实施

616　在这一章中，我们将以完成一篇学期论文为重点，讨论一项成功的经验实证分析的构成要素。在此，除了帮助你回忆全书中出现过的重要内容以外，还会强调在应用研究中反复出现的重要主题。为了激发读者的想像力，我们还对一些题目提供建议，并给出一些经济研究的材料和数据供读者参考。

19.1　问题的提出

提出一个非常确切的问题，其重要性是不容忽视的。如果没有明确的分析目标，那么你的研究将无从下手。如果因为丰富的数据集具备了广泛的适用性，你就企图在想法尚未成熟之时开始搜集资料，其结果往往适得其反。如果没有对你的假设和你将要估计的该类模型进行细致的公式化表述，那么你很有可能会忘记搜集某些重要变量的信息，或是从错误的样本空间中取样，甚至会在搜集数据时对应错了时间区间。

当然，这并不是说你要凭空臆造一个问题，尤其是在做学期论文的时候，更不可能去建造空中楼阁。因此在选择题目的时候，你必须确信现有的

数据来源能够让你在指定的时间里回答你的问题。

在选题时，你必须明确，你对经济学或者其他社会科学的哪一个领域感兴趣。举例来说，在完成了“劳动经济学”的课程后，你可能会发现其中的理论能够被实践所检验，或是这些理论与一些相关政策存在着联系。劳动经济学家不断地发现能够解释工资差异的新变量，这包括高中阶段的教学质量（Card and Krueger，1992；and Betts，1995），数理基础的课程量（Levine and Zimmerman，1995），以及学生的身体特征（Hamermesh and Biddle，1994；Averett and Korenman，1996；Biddle and Hamermesh，1998）。而国家或地方的公共财政研究人员则致力于研究当地的经济活动是如何依赖于经济政策变量的，这些变量包括：财产税、销售税、服务的水平和质量（如学校、消防队和警察局）等等（White，1986；Papke，1987；Bartik，1991；Netzer，1992）。研究教育问题的经济学家则对以下三个问题颇为关注（Hanushek，1986）：支出如何影响求学行为，就读某类学校是否会提高受教育者的能力，以及如何确定影响私人学校选址的因素（Downes and Greenstein，1996）。宏观经济学家对各种各样的总体时间序列之间的关系兴趣十足，如国民生产总值的增长与固定资产投资的增长（如机器设备的增长）之间的关系，或是税收对利率的影响（Peek，1982）。

评估的模型通常都具备可描述性，这是非常有道理的。举例来说，财产
617 评估者利用Hedonic定价模型（见例4.8）对某家庭最近尚未卖出的房屋价值进行评估。它描述了房屋的价格对它的特性（大小、卧室的数量、浴室的数量等）的回归模型。若以此作为学期论文就不新奇了，因为我们不可能从中学到更多有新意的东西，而且这些分析也没有什么明显的政策内涵。可是如果把邻里犯罪率作为一个解释变量加进来分析，就能够知道犯罪率在确定房屋价格时是一个非常重要的因素。这在评估犯罪成本时会有帮助。

一些关系式的估计大多是利用了描述性的宏观经济数据。例如，一个总量储蓄函数能用来判断总量边际储蓄倾向和储蓄对资产的回报（如利率）的反应。如果把时间序列数据应用在一个曾经经历过政治动乱的国家，并确定其在政治不稳定时期储蓄率是否会下降，这样的分析将会更有趣。

一旦你确定了一个研究的领域，对于所选的题目就可能有许多的方法来为论文定位。*Journal of Economic Literature*（JEL）有一套细致的分类体系。每篇论文都有一组确认编号，从而将其归于经济学的某一子领域中。JEL还网罗了在其他各类期刊中发表的文章的列表，按其题目来组织，甚至包括文章的摘要。

因特网（Internet）服务使得搜寻各种题目的已发表论文更为方便。例如，*Econlist* 就为许多大学所订阅。使用者通过 *Econlist*，按照作者名字、主题、题目中的关键词等方式便可以广泛地搜寻几乎所有的经济学期刊。Social Science Citation Index用在寻找与社会科学的广泛领域相关的论文时非常有用，尤其包括那些很受欢迎并经常被其他著作所引用的优秀论文。

在构思一个题目时，对以下几个问题要做到心里有数。第一，要使一个问题让人感兴趣，并不需要它具有广泛的政策内涵，相反，它可以只是当地

的兴趣所在。例如在大学里，生活在社团中间会使学业成绩提高还是降低，也许是你关心的问题。这一问题或许能或许不能令校外的人感兴趣，但至少会引起校内部分人的关注。另一方面，你也许只是从当地的利益出发分析问题，最后却引起广泛的影响。比方说确定什么因素会影响大学校园的酗酒现象，以及有哪些学校规章可以对其进行制约等等。

第二，利用美国经济的标准宏观经济总量数据来进行真正有创造性的研究是非常困难的，尤其对于一篇要在半个学期或一个学期之内完成的论文来说更是如此。比方说，货币增长和政府支出增长等因素是否会影响经济增长，是专业的宏观经济学家一直以来致力研究的问题。利用现有的信息能否 618 系统地预测股票或其他资产的回报，这一问题显然已经被研究得非常仔细。然而，这并不意味着你应该回避对宏观或经验性金融模型的估计，因为仅仅增加一些更为新近的数据也可以使我们的争论更有建设性。另外，有时你会发现对经济总量和金融回报有重要作用的新变量，这样的发现往往令人非常激动。

问题是，再多用几年的数据为美国经济或一些更大的经济社会估计一个标准的菲利普斯曲线或总量消费函数，像这样的练习，并不可能为我们带来更多新的理解，尽管它们对学生来说会有些启发。然而，你可以利用一个较小国家的数据来估计静态或动态的菲利普斯曲线，或是检验有效市场假说等等。

在非宏观层面上，同样有许多被广泛研究的问题。劳动经济学家发表了许多关于教育回报率的论文。这个问题因为很重要，所以还在被研究，况且新的数据集，还有新的计量经济学方法也在发展之中。举例来说，正如我们在第 9 章中看到的，对不可观测的能力来说，有些数据集就比其他数据集提供更好的代理变量。（比较 WAGE1.RAW 和 WAGE2.RAW。）另外，我们可能得到综列数据或者从一个自然实验得到数据——见第 13 章——使得我们得以从一个不同的方面去思考一个老问题。

再举一例，犯罪学家对研究不同的法律对犯罪的影响感兴趣。死刑是否有威慑作用，长久以来成为一个争论的话题。类似地，经济学家对税收是否能减少烟酒的消费量感兴趣（一如既往在其他条件不变的意义下）。随着我们掌握越来越多年的州一级数据，一个更为丰富的综列数据集就会产生，它能够帮助我们更好地回答大部分政策问题。另外，有一些相当新近的反犯罪革新——例如社区管辖的出现——它们的有效性如何可以经验性地评估出来。

在把你的问题作公式化表述的时候，跟你的同学、导师和朋友讨论你的想法会有所帮助。你应该能够说服别人相信你对问题的解答是值得关注的。（当然，你是否能够有说服力地回答你的问题是另一回事，但你需要从一个值得关注的问题开始。）如果有人询问你的论文，而你的回答是："我正在进行关于犯罪的论文"或是"我的论文是关于利率的"，这极有可能说明你只确定了一个总的范围，却没有公式化表述出一个真正的问题。你应该能够说出一些"我正在研究美国的社区管辖对城市犯罪率的影响"或者"我正着眼

于巴西的通货膨胀的反复无常是如何影响短期利率的”之类的话。

19.2 文献回顾

所有的论文，即使非常短，也应该包括相关的文献综述。几乎没有人会试图进行尚无发表先例的经验项目。如果你通过期刊或**网络搜寻服务**（online search services），例如 *Econlist* 搜寻一个题目，那么你所做的正是文献检阅。619 如果你自行选题——例如研究你们学校药用量对学业表现的影响——那么可能工作起来要辛苦一些。但网络搜寻服务使这一过程简化，例如你可以通过关键词、题目中的字或者作者等等进行搜索。然后你就可以通过论文的摘要得知它们与你自己的研究有多少关系。

在进行你的文献检阅时，你应该考虑到利用少量关键词可能搜寻不到的相关题目。例如，如果你在研究吸毒量对工资或平均成绩的影响，或许你应该查阅一些关于饮酒量对这些因素影响的著作。了解如何进行彻底的文献搜寻是一种有待学会的技巧，但如果你在行动前多加思考，可能会减少不少麻烦。

如何将文献检阅包含在论文中，不同的研究人员有不同的做法。有些人喜欢开辟独立的章节，叫做“文献检阅”，而其他人则倾向于将文献检阅作为一个部分包括在序言中。这主要视个人偏好而定，尽管大量的文献检阅是值得占有它独立的章节的。如果学期论文是课程的重点——例如在一个高级专题或是高等计量经济学课程中——你的文献检阅就有可能相当长。一篇入门课的期末论文往往较短，相应的文献检阅也要简洁一些。

19.3 数据的收集

确定适当的数据集

为学期论文搜集数据可能是富有教育意义的、令人激动的，有时也可能是相当叫人沮丧的一件事。首先你必须确定用以回答你所设定的问题的数据 620 类型。正如我们在序言中讨论过并贯穿于本书的观点，数据集以多种形式出现。最常见的类型是横截面、时间序列、混合截面和综列数据集。

有些问题可以由我们介绍过的任何一种数据结构进行描述。例如，在研究是否更强的执法力度能够降低犯罪率时，我们可以利用一些城市的横截面数据，或者某个给定城市的时间序列数据，也可以选择一些城市的综列数据

集——它包含相同几个城市两年或者多年的数据。

确定搜集何种数据通常取决于该分析的性质。为了回答个人或家庭层面的问题，我们通常只需要找到单一横截面的数据；它们往往由调查取得。接着，我们要自问，是否能够获得一个足够丰富的数据集来进行在其他条件不变下的分析。举例来说，假设我们想知道是否通过个人退休账户（IRAs）来储蓄的家庭——这具有某些税收优势——会有较少的非 IRA 储蓄。换言之，是否 IRA 储蓄不免会把其他形式的储蓄排挤出去？有些数据集，例如消费者财务调查，包括了每年不同家庭样本的不同储蓄种类的信息。利用这样的数据集会产生一些问题，其中或许最重要的一个是，是否有足够多的控制变量——包括收入、人口统计特性和储蓄偏好的代理变量——来进行在其他条件不变下合理的分析。如果这些是我们可以获得的惟一一种数据集，就必须尽我们所能来处理它们。

在处理关于公司、城市、州县等的横截面数据时，相同的问题也会产生。在大多数情况下，我们能否利用单一横截面的数据进行在其他条件不变下的分析，这一点并不明显。例如，关于执法力度对犯罪率的作用的任何研究都必须认识到执法经费的内生性。如果利用标准的回归方法，那么无论我们有多少控制变量，或许都很难完成一个令人信服的在其他条件不变下的分析。（参见 19.4 节更多的讨论。）

一旦阅读了关于综列数据方法的较高深的章节后，你就会发现，具备两个或多个不同时点上的相同的横截面单元的数据能够让我们控制好不随时间而改变的无法观测到的效应，而正是这些效应经常把我们在单一横截面上所做的回归混杂起来。个人或家庭的综列数据集相对来说较难获得——尽管还是有一些例子存在，如动态收入的综列研究——但它们可以用于非常令人信服的渠道。公司的综列数据也是存在的。例如，Compustat 和证券价格研究中心（CRSP）对各公司进行着大规模的财务信息的综列数据信息的收集。对更大的单元，如学校、城市、郡县和州来说，要获得综列数据则更为容易，因为这些单元不会随着时间而消失，而且政府机构会负责每年收集这些相同变量的信息。例如，联邦调查局会收集并报告每个城市的犯罪率的详细信息。本章附录列出了一些数据的来源。

数据以多种形式出现。一些数据集，特别是历史数据，通常仅以印刷资料的形式出现。如果数据集不大，把来源于印刷资料的数据输入在操作上是便利的。有时候，有些文章连同小的数据集一起发表——特别是时间序列应用。把这些资源应用于经验研究中，也许还得用近年来的数据加以补充。

许多数据集可以在计算机磁盘或磁带中找到。前一种特别易于操作。现在，很大的数据集已经可以存放在磁盘中了。各种政府机构会出售数据磁盘，私人企业也不例外。论文作者通常愿意用磁盘的形式提供他们的数据集。

越来越多的数据集可以在互联网上找到。网络是**网上数据库**（on-line data bases）的丰富来源。无数包括经济和相关数据集的网站被建立起来。有一些网站包括经济学家所关注的数据集的链接；本章的附录列举了一些这样

的网站。总的来说，用因特网搜寻数据资源相当简单而且在将来会更加便捷。

输入并储存你的数据

一旦你确定了数据类型并找到了数据来源，就必须把数据转变为可操作的形式。如果数据在磁盘里，那么它们已经具备了一定的形式，而且是有望具备广泛用途的一种形式。用磁盘形式获取数据最灵活的方法是将其作为标准的**文本**（ASCII）文件（text file）。

621 所有的统计和计量经济学软件包都可以以这种方式存储原始数据。一般说来，只要该文本文件构架合理，就可以被直接读入一个计量经济学的软件。我们在全书中所使用的数据文件提供了横截面、时间序列、混合截面和综列数据通常的存储方式的例子。作为一般性原则，数据应该具备表格形式，每次观测占一个不同的行；而数据集的每一列则代表不同的变量。偶尔你会遇到以列代表观测次数而行代表不同变量的方式存储的数据集。这并不是理想的方式，但大部分软件包允许以这种形式读取数据，然后把它改过来。所以自然地，了解数据读入计量经济学软件之前如何组织是非常关键的。

对时间序列数据集来说，只有一种实用的方式来进行数据的输入和存储，也就是说，以时间为序，最早的时间段列为第一次观测，最近的时间段列为最后一次观测。把带有年份或（如有必要）带有季度或月份的变量包括进来通常会有所帮助。它有助于对今后模型有所改变时的估计，包括考虑到不同时间区间的季节性变化和间断。对一段时间的混合截面来说，通常最好是把最早一年的截面放在第一个观测区里，接着是第二年的截面，然后依次下去。（参见 FERTIL1.RAW。）这种安排并不关键，但把注明年份的变量附在每一次观测上是非常重要的。

对综列数据来说，正如我们在 13.5 节讨论过的，如果所有年份的横截面观测都是相邻的且以时间为序，就是最理想的。有了这样的顺序，我们就可以利用第 13 章和第 14 章的所有综列数据的方法。对于综列数据，让每个横截面单元具备一个独特的标识符和一个年变量是非常重要的。

如果你获得印刷资料形式的数据，就可以有多种选择将其输入电脑。首先，你可以利用标准**文本编辑器**（text editor）创建一个文本文件。（这就是一些包括在文本中的原始数据集最初创建的方法。）一般的要求是每一行开始一次新的观测，并且每行的变量有相同的顺序——特别地，每行的条目数要相等——而且数值与数值之间至少需要由一个空格分开。有时候，用一个不同的分隔符，如逗号，可能会更好，但这取决于你所使用的软件。如果你缺掉了某些变量的某些次观测，你就必须决定如何将其表示出来；仅仅留下一个空格一般是不行的。许多回归软件包接受以句号为缺落数值的标志。有些人倾向于用一个数字——也许是对所关注的变量来说是一个不可能的值

——来表示缺落数值。如果你不是非常细心的人，这样做就非常危险；我们后面要进一步讨论这个问题。

如果你有非数值数据——比方说，你想把一个学院样本的名字或一些城市的名字包括进来——那么你需要查询一下你将要使用的计量经济学软件包以确定输入这种变量的最佳方法（通常叫做字符串）。习惯上字符串都被放置在双引号或单引号中间。有的文本文件遵循严格的形式，即通常用一个小程序来读取文本文件。不过你还是要查询你的计算机软件包以获得详细的信息。

另一种普遍适用的选择是利用**总分析表**（spreadsheet）来输入你的数据，比方说Excel。它与文本文件相比有一些优点。第一，因为每个变量的每次观测是一个单元，所以数字不太可能撞在一起。（如果你忘记在数字之间输入空格，这种情形在文本文件中就会出现。）第二，总分析表允许对数据进 622 行操作，例如进行归类和计算平均值等。如果你所使用的软件包能进行精密的数据管理，那么这第二个优点就不是那么重要了；许多软件包，包括E-views和Stata，都属于这一类。如果你利用总分析表进行原始的数据输入，那么通常必须把数据以你的计量经济学软件包可读的形式报告出来。通常这是非常明了的，因为从总分析表输出到文本文件可使用多种格式。

第三种替代方法是直接把数据输入你的计量经济学软件包。这种方法就不需要文本编辑器或是总分析表，但是更为笨拙，因为你不能在不同的观测上通过自由移动来进行更改或添加。

从因特网上下载数据以多种形式出现，通常数据是文本文件的形式，但不同的规则可以处理不同的变量；对综列数据集来说，如何将数据排序，不同的规则会有所不同。一些因特网数据集以总分析表的形式出现，这种情况下你就必须用合适的总分析表来读取它们。

检查、清理、总结你的数据

在经验分析中，熟悉你将要使用的数据集尤其重要。如果你自己输入数据，那么你就必须完全了解它的内容。但如果你从外界来源获取数据，就仍然需要花时间了解它的结构和管理。即使是广泛使用且大量储存的数据也会有缺陷，如果你使用来自某论文作者的数据集，你就必须意识到数据集的构筑方式可能会被忽略。

前面我们回顾了不同数据集存储的标准方法。你还需要知道缺失数值是如何标识的。较可取的方法是用一个非数字符号，如句号，标明缺失数值。如果是利用诸如“999”或者“−1”这些数字来作为缺失数值的标识，那么在利用这些观测进行统计计算时就必须非常小心。你的计量经济学软件包或许不知道某一数字实际上表示一个缺失数值：很有可能这样一次观测会被当做有效而使用，从而产生极其错误的结果。最好的办法是把所有代表缺失数值的数字符号改为其他不会与实数数据混淆的符号（如句号）。

你还必须知道数据集中变量的性质。哪些是二进制数？那些是序数（如信用评级）？变量的测量单位是什么？比方说，货币价值是用美元、千美元、百万美元还是其他什么单位测量的？变量是表示以百分比还是比例测量的比率？——例如学校的退学率、通货膨胀率、参加工会率或者利率。

特别是对于时间序列数据来说，了解货币价值是名义的（当前的）还是实际的（恒常的）美元非常关键。如果这个价值是实际值，那么基年或基期又是什么？

如果你从一个作者那里得到一个数据集，那么有些变量已经以一定的方法进行了变换。比方说，有时候只有变量的对数形式（如工资或薪水）出现在数据集中。

在数据集中检测错误对保持任何数据分析的完整性都是必要的。在分析中找出所有的或者至少是最显著的变量的最小值、最大值、均值和标准差通常都是有用的。举例来说，如果你发现你的样本中教育的最小值是 -99，你就会知道教育的输入值中至少有一个要被设为缺失数值。经过进一步检查，623 如果教育水平有数次观测值为 -99，那么你可以信心十足地说，你发现了教育的缺失数值的标志。举另一例来看，如果你发现一个城市样本的谋杀罪平均定罪率为 0.632，你就知道这个定罪率是以比率而不是百分比测量的。那么，如果最大值大于 1，那么这有可能是一个印刷错误。（发现数据集中大部分比率变量以百分比输入，一部分以比率输入，或者反过来，这些情况并不罕见。这样的数据标志错误难以发现，但进行这样的尝试是很重要的。）

我们在使用时间序列数据时同样要小心。如果运用月份或季度数据，我们就必须知道哪些变量，如果有的话，是经过了季节性调整的。改变数据同样需要格外小心。假使我们有每月的数据集，而且想要从中产生从一个月到下一个月的改变量。为达到这一目的，我们必须确定数据是按日历从最早的时期到最晚的时期排序的。如果出于某些原因情况并非如此，相减的结果将是无意义的。为确保数据正确排列，运用一个时间标示变量会有所帮助。对于年份数据，知道年份就足够了，但我们还必须知道年份是以 4 位数还是 2 位数输入的（比方说，1998 或 98）。同样，有了月份或季度数据，拥有一个或多个标示月份或季度的变量将有所帮助。对于月份数据，我们可以拥有一组虚拟变量（11 个或 12 个）或者用一个指明月份的变量（1～12 或是一个字符串变量，如 jan，feb 等）。

无论用不用年、月或季度标示变量，我们都可以在所有计量经济学软件包中轻易地构建时间趋势。如果标示出月份或季度，那么构建季节虚拟变量就很简单；至少，我们需要知道第一次观测的月份或季度。

操作综列数据更具挑战性。在第 13 章中，作为控制不可测影响的一般方法，我们讨论了对差分数据的混合 OLS。在建立差分数据的时候，我们应该小心，不要产生幻象观测。假使我们拥有 1992—1997 年城市的平衡综列数据。即使数据在每一个横截面单元上都按时间先后排序——这是在开始之前首先要做的事情——一个不经心的差分也会对样本中除第一个之外的所有城市产生一个 1992 年的观测值。这一观测值将是 1992 年城市 i 的值减去

1997 年城市 $i-1$ 的值；这显然是愚蠢的。因此，我们必须保证，对所有被取差分的变量来说，1992 年都是缺失了的。

对非平衡综列数据来说，情况会更曲折，因为没有一个命令是对所有横截面单元都适用的。通常对非平衡综列数据使用固定影响估计会更简便。

19.4 计量经济学分析

本书集中介绍计量经济学分析，但在本节中我们不准备对计量经济学方法作一复习。不过，对于在经验分析中需要考虑的问题，我们会提供一些一般性的指导。

624 就像早先讨论过的，在确定了题目之后，必须选出一个合适的数据集。假定这也已经完成了，我们就必须接着决定合适的计量经济学方法。

如果你的课程着重于多元线性回归模型的普通最小二乘估计，利用的数据是横截面或者时间序列，那么对你来说计量经济学方法在很大程度上就已经确定了。这不一定有坏处，因为 OLS 仍然是使用最广泛的方法。当然，你还必须确定，是否需要 OLS 的任何一种变形——例如加权最小二乘或时间序列回归中的序列相关的校正。

为说明 OLS，你还必须给出一个令人信服的例子以说明 OLS 的关键假定满足你的模型。如同我们详细讨论过的，第一个问题是，误差项是否与解释变量不相关。较理想的情况是，你能够控制足够多的其他变量来假定留在误差项中的因素与回归元不相关。特别是在处理个人、家庭或者公司这一层次的横截面数据时，自我选择问题——我们在第 7 章和第 15 章中讨论过的——通常是需要考虑的。例如，在 19.3 节中的 IRA 的例子中，或许对储蓄有不可观测偏好的家庭也正是那些开立了 IRA 的。你还应该能够证明其他潜在的内生因素——也就是，测量误差和联立性——不是严重的问题。

在设定你的模型的时候，你同样需要确定采取何种函数形式。某些变量是否应该以对数形式出现？（在计量经济学应用中，答案通常是肯定的。）某些变量是否应该为水平值或平方值描绘可能正在减少的影响？定性因素应该怎样出现？对不同的属性和群体，仅用二值（虚拟）变量够吗？或者，需要考虑它们与数量变量的交互作用吗？（见第 7 章的详细情况。）

对横截面分析来说，其次但同样非常重要的问题是，异方差是否存在。在第 8 章中，我们说明了如何解决这一问题的方法。最简单的办法是计算对异方差稳健统计量。

正如我们在第 10 章、第 11 章和第 12 章中强调过的，对时间序列的应用需要格外小心。对方程的估计应使用水平值吗？如果利用水平值，是否需要时间趋势变量呢？用数据的差分是否更合适？如果是月份数据或季度数据，是否应该考虑季节因素？如果你考虑动态——比方说，分布滞后动态——有多少滞后因素应该被包括进去？你应该从一些基于直觉或常识的滞后

因素开始，但这最终还是一个经验问题。

如果你的模型有潜在的设定错误，比方说遗漏变量，而且利用 OLS，那么你应该尝试一些我们在第 3 章和第 5 章中讨论过的误设分析。基于合理假定，你能够确定估计量存在着哪一方向的偏误吗？

如果你已经学习了工具变量的方法，你就知道可以用它来解决各种形式的内生问题，包括遗漏变量（第 15 章）、变量误差（第 15 章）和联立性（第 16 章）。自然地，你需要仔细想想你所考虑的工具变量是否可能奏效。

经验社会科学方面的优秀论文应包含**敏感度分析**（sensitivity analysis）。宽泛地说，这意味着你首先估计一个初始模型，然后用一些看似合理的方法修改它。理想的情况是重要的结论不致改变。例如，如果你把酒精消费量的 625 一个度量作为解释变量（如在平均成绩的方程中），或者用一个表示酒精用量的虚拟变量来替代定量的度量，这两种做法能够得到性质上相似的结果吗？如果表示用量的二值变量是显著的，而表示酒精量的数量变量不是，那么酒精用量就可能反映出某些影响着 *GPA* 且与酒精用量相关的不可观测的属性，但这需要根据具体情况进行考虑。

如果某些观测值与样本群体非常不同——比方说，在样本中，有几个公司比其他公司大得多——那么如果把这些观测值从估计中排除出去，你的结果会有很大改变吗？如果是，你就可能不得不改变函数的形式来考虑这些观测，或者证明它们从属于完全不同的一个模型。异常值问题曾在第 9 章中进行了讨论。

对综列数据的利用为我们提出了更多的计量经济学问题。假使你已经搜集了两个时期的数据，无须借助工具变量，至少也有 4 种方法来利用这两个时期的综列数据。你可以在标准 OLS 分析中混合这两个时期，就像在第 13 章中讨论过的一样。虽然相对于一个单一截面来说，这样做或许可以扩大样本容量，但它并没有控制不随时间而变化的不可观测变量。另外，由于观测不到的影响，这样一个方程中的误差几乎总是序列相关的。只要在所有时间区间中给定解释变量的值之下，观测不到的影响有零均值，用随机效应估计法就可以纠正序列相关问题并且产生渐进有效的估计量。

另一种可能的方法是在第二年的方程中加入滞后的因变量。在第 9 章中，我们把它作为至少可以缓解缺落变量问题的一种方法给出，就像我们在任何情况下都要保持因变量的最初结果固定不变那样。正如我们在第 13 章中讨论过的，这通常会导致与取数据差分相似的结果。

当我们拥有更多年的综列数据时，除了老办法，还有一种新的选择。我们可以利用固定效应变换来消去不可观测因素的影响。（只有两年的数据，变换效果和取差分一样。）在第 15 章中，我们说明了工具变量技术如何与综列数据变换结合，以进一步放宽外生性假定。作为一般性原则，同时运用几种合理的计量经济学方法并比较它们的结果是一种很好的想法，这通常能够帮助我们确定诸假定中的哪一个可能是错误的。

设计题目、设想模型、搜集数据并运用计量经济学方法，即使在这整个过程中你都非常谨慎，你还是很有可能——至少在某些时候——得到令人迷

感的结果。当这种情况发生时，很自然的想法是尝试不同的模型、不同的估计方法，或者可能不同的数据子集，直到结果与预期的更加一致，实际上所有的应用研究人员在找到“最佳”模型之前会寻找各种不同的模型。不幸的 626 是，**数据开采**（*data mining*）的实践破坏了我们在计量经济学分析中作出的假定。OLS 无偏性和其他估计量的结果以及我们为假设检验而推导的 t 分布和 F 分布，都假定了我们按照总体模型来观测一个样本并对该模型进行了一次估计。估计一个与初始模型相异的模型破坏了这一假定，因为我们在模型设定的搜索时利用了相同的数据集。实际上，通过利用这些数据来重新设定我们的模型，我们利用了检验的结果。从不同的模型设定中得到的估计值和检验值就不是互相独立的。

一些设定搜索已经被编制为标准软件包中的程序。最普通的一种是逐步回归，也就是在多元回归分析中运用了解释变量的不同组合以试图获得最优的模型。利用逐步回归可以有不同的方法，在这里我们不想评论它们。一般的思路是，或者从一个大的模型开始，然后保留在某一显著水平之下的变量，或者是从一个简单模型开始，然后加入 p 值显著的变量。有时候也用 F 检验来检验一组变量。不幸的是，最终的模型通常取决于变量被丢弃或添加的顺序。[参见德雷珀和史密斯（Draper and Smith，1981）更多的内容。] 另外，这是数据开采的一种严格形式，而且在最终的模型中也难以解释 t 和 F 统计量。有人也许会说，逐步回归只是把研究人员在寻找不同的模型时所做的事自动化了。然而，在大部分的应用中，一个或两个解释变量是我们的主要关注所在，目标是了解这些变量的系数有多强，从而决定应该增加或丢弃其他变量，或者改变其函数形式。

原则上，把数据开采的影响包括到我们的统计推导中是可能的，但在实际操作中却是非常困难且极少完成的，特别是在复杂的经验工作中更是如此。[参见利墨（Leamer，1983）对这一问题充满魅力的描述。] 不妨直至找到一个显著的结果，才开始寻找大量的模型或估计方法，然后仅报告这个结果，这样做我们也许可以把数据开采的弊端最小化。如果一个变量仅在诸多估计模型的一小部分中统计上显著，那么这个变量极有可能在总体中不起作用。

19.5 经验论文的写作

运用计量经济学分析来写作一篇论文是非常富有挑战性的，同时回报也是丰厚的。一篇成功的论文包括仔细而令人信服的数据分析与优秀的解释和展示相结合。因此，你必须对你的题目有很好的把握，对计量经济学方法有很好的理解，并且具备扎实的写作技巧。可是如果你发现写作经验论文非常困难，也不要灰心；大部分专业研究人员花费了多年的时间来学习如何巧妙地进行经验分析并以令人信服的形式写出其结果。

虽然写作风格各不相同，但许多论文还是遵从相同的基本轮廓的。下面的段落包括了关于章节标题的看法，并解释了每一节应该涵盖的内容。这只是一些建议，不一定要严格遵守。在期末论文中，每节都应以数字进行标识，通常从引言标识为1开始。

引言

引言阐述研究的基本目标并解释其重要性。它一般包括一个文献检阅，表明有哪些工作已经完成以及过去的工作可以怎样改进。(正如我们在19.2节中讨论过的，篇幅较长的文献检阅应自成一个独立的章节。)展示一些简单的统计表或者图表来揭示一个看似矛盾的关系，是介绍论文主题的有用方法。举例来说，假使你正在写一篇关于某个发展中国家影响生育率的因素的论文，并把重点放在妇女的受教育水平上。介绍这一主题的富于感染力的方法是创建一张表或一幅图，来显示（比方说）一段时间内生育率的下降，并简述你希望如何研究导致这一下降的因素。在这一点上，或许你已经知道，在其他条件不变的情况下，受教育水平越高的女性生育的孩子越少，而随着时间的消逝，平均教育水平提高了。 621

大部分研究人员喜欢在引言中总结其论文的发现，这是吸引读者注意力的有效方法。例如，在一个30学时的学期课中缺掉了10个小时的课程，其影响如何？也许你会阐明你对这一影响的最佳估计值大约是1/2的积点分。但总结不应该太深入，因为得到这些估计值的方法和数据都还没有介绍过。

概念（或理论）框架

在这一节中，要描述你所提问题的主要解决办法。它可以是正式的经济学原理，但在许多情况下，它只是在回答你的问题时出现的概念性问题的直觉性讨论。

举例来说，假使你正在研究经济机会和惩罚的严厉性对犯罪行为的影响。解释参与犯罪的一个方法是，在给定合法和非法的工资率、测量犯罪活动的概率以及惩罚的严厉性的变量的情况下，说明个人选择花费在合法及非法活动上的时间效用最大化的问题。这种方法的用处在于，它示意你哪些变量应包括在经验分析中；对这些变量应如何出现在模型中给出一般性指导。

通常没有必要写出一个经济理论来。对计量经济学政策分析来说，往往常识就足以设定一个模型。例如，假使你对估计“有未成年儿童家庭的救济(AFDC)”的参与对儿童在校表现的影响感兴趣。AFDC提供收入补贴，但参与其中同时使获得医药救助和其他福利更为容易。进行这样一个分析的艰难之处在于确定所要控制的一组变量。在这个例子中，我们可以控制家庭收入（包括AFDC和任何其他福利收入）、母亲的受教育水平、家庭是否居住

于城区以及其他一些变量。然后引进一个 AFDC 参与变量就（可望）测出参与 AFDC 的非收入福利。用关于应控制哪些变量的讨论以及对参与 AFDC 如何提高在校表现的描述去替代正式的经济学原理。

计量经济学模型和估计方法

开辟一个章节来描述一下你所估计的并且将要在论文的结果章节中展示的那些方程式，很有好处。它可以帮助你整理思路，从而让你明确关键的
628 解释变量是什么以及应该控制其他哪些因素。写出一些包含误差项的方程式好让你考虑诸如 OLS 的方法是否合适的问题。

对一个模型和一个估计方法的区别应该在这一节中指出。一个模型代表了一个总体关系（广泛定义使之包括时间序列方程）。例如，我们应该写出

$$colGPA = \beta_0 + \beta_1 alcohol + \beta_2 hsGPA + \beta_3 SAT + \beta_4 female + u \quad (19.1)$$

来描述大学 *GPA* 和饮酒量之间的关系，同时控制方程中的一些其他变量。假定这一方程代表一个总体，比方说一个大学的所有本科生。β_j 或 *colGPA* 上没有“帽”（ˆ），因为这是一个模型，而不是估计方程。我们没有写出 β_j 的数值，因为我们不知道（而且永远也不会知道）这些数值。以后我们会估计它们。在这一节中，不要提前进行经验结果的展示。换句话说，不要从一个一般模型一开始就说你删去一些变量，因为它们被证明是不显著的。这样的讨论应该留到结果章节中。

联系城市汽车偷窃率和失业率（以及其他控制变量）的一个时间序列模型有如

$$\begin{aligned} thefts_t = {} & \beta_0 + \beta_1 unem_t + \beta_2 unem_{t-1} + \beta_3 cars_t \\ & + \beta_4 convrate_t + \beta_5 convrate_{t-1} + u_t \end{aligned} \quad (19.2)$$

式中，t 下标对强调方程中的动态性质很有用（在这个例子中，要考虑失业率和机动车偷窃定罪率有滞后影响）。

在说明了一个或几个模型之后，讨论估计方法就显得合适了。在大多数情况下用到的是 OLS，但是，比方说在时间序列方程中，你也许会用适当的 GLS 来进行序列相关校正（见第 12 章）。但是，估计模型的方法与模型本身要明显区别开来。说诸如“一个 OLS 模型”这样的话是没什么意义的。普通最小二乘是一种估计方法，加权最小二乘法、科克伦-奥克特方法等等也是。通常对任何一个模型有许多估计方法，你应该解释为什么你所选择的方法是有道理的。

从一个基础的经济学模型得到一个可估计的计量经济学模型，在其中用到的任何假定都要进行清晰的讨论。例如，在 19.1 节中提到过的高中质量的例子，如何测量学校质量的问题是分析的中心部分。是应该建立在平均 *SAT* 分数上、入读学生的毕业比例上、学生与老师的人数比例上还是老师的平均教育水平上？抑或是这些因素的组合或者其他可能的度量？

无论理论模型是否已得到展示，我们通常都要对函数形式作出假定。如你所知，常弹性和半常弹性模型很有吸引力，因为系数易于解释（作为百分比）。如何选择函数形式没有硬性的原则，但6.2节中讨论过的准则在实践中却很起作用。你不需要对函数形式进行长篇讨论，但略为提及你要估计的
629 是弹性还是半弹性模型是有帮助的。举例来说，如果你要估计某变量对工资或薪水的影响，因变量几乎可以肯定是采用对数形式的，而且你也许从一开始就把它包括到了每一个方程中。你不需要阐述所有的，甚至只是大部分的，你将在结果章节中报告的函数形式的变化。

通常经验经济学中使用的数据是以城市或郡县为单元的。举例来说，假使总体是由小型到中型城市构成的，你想要检验这样一个假设：拥有一只小规模棒球队能导致一个城市的离婚率下降吗？在这个例子中，你必须考虑越大的城市离婚人数越多这样一个事实。考虑城市大小的一个方法是以城市或成年人为单位来度量离婚数量。因此，一个合理的模型是

$$\begin{aligned}\log(div/pop) &= \beta_0 + \beta_1 mlb + \beta_2 perCath \\ &\quad + \beta_3\log(inc/pop) + \text{其他因数} \qquad (19.3)\end{aligned}$$

式中，*mlb* 为一个虚拟变量，如果该城市有一支小型棒球队，变量就为1；*perCath* 为人口中天主教徒的百分数（因此它若为34.6则表示34.6%）。注意 *div/pop* 是一个离婚比率，它一般来说比离婚的绝对数易于解释。

控制总体的另一方法是估计模型

$$\begin{aligned}\log(div) &= \gamma_0 + \gamma_1 mlb + \gamma_2 perCath + \gamma_3\log(inc) \\ &\quad + \gamma_4\log(pop) + \text{其他因数} \qquad (19.4)\end{aligned}$$

保持人口、天主教徒百分率、收入和其他任何因素不变，当我们关注的参数 γ_1 被乘以100之后，就给出了离婚率的百分比差异。在方程（19.3）中，β_1 测量了一支小棒球队对 *div/pop* 的百分比影响，当离婚数量或人口二者之一发生变化时，它都会改变。利用事实：$\log(div/pop) = \log(div) - \log(pop)$ 和 $\log(inc/pop) = \log(inc) - \log(pop)$，我们可以把式（19.3）改写为

$$\begin{aligned}\log(div) &= \beta_0 + \beta_1 mlb + \beta_2 percath + \beta_3\log(inc) \\ &\quad + (1-\beta_3)\log(pop) + \text{其他因数}\end{aligned}$$

它说明式（19.3）是当 $\gamma_4 = (1-\beta_3)$ 且 $\gamma_j = \beta_j$（$j = 0, 1, 2, 3$）时式（19.4）的一个特例。等价地，式（19.4）相当于是把 $\log(pop)$ 作为额外的解释变量加入到式（19.3）。这使得检验单一个人口（变量）对离婚率的影响变得简单。

如果你利用一个更先进的估计方法，例如二步最小二乘，你需要就你的选择作出解释。如果你用2SLS，你需要就你对一个（或多个）内生解释变量所做的IV选择的有效性进行详细的讨论。正如我们在第15章中提到过的，一个变量要成为一个好的IV需要两个条件。首先，它必须被关注方程（结构方程）所略去，并外生于该方程。这是我们必须假定的。第二，它必

须与内生解释变量有某种偏相关。我们可以检验这一点。比方说，在方程(19.1)中，作为饮酒量的 IV，你也许会用来表示一个学生是否住在公寓的
630 一个二值变量。这要求居住环境对 $colGPA$ 没有直接影响——使得它可被方程（19.1）所忽略——而且它与 u 中影响 $colGPA$ 的不可测因素不相关。我们同样还需要通过 $alcohol$ 对 $dorm$，$hsGPA$，SAT 和 $female$ 的回归，来证明 $dorm$ 与 $alcohol$ 偏相关。（见第 15 章的详细内容。）

在利用综列数据时，你或许会考虑忽略变量（或者忽略异质性）的问题。同样地，写出一两个方程，这个问题就易于描述了。事实上，在说明一段时间如何对方程求差分以消除不随时间而变的不可观测量时，这样做就很有用；它给出了一个可由 OLS 估计出来的方程。或者，如果你利用的是固定效应估计，你只需要简单说明即可。

作为一个简单的例子，假使你正在检验是否较高的郡县税率会减少经济活动（如用人均产出来度量的）。假使对 1982 年、1987 年和 1992 年，该模型为

$$\log(manuf_{it}) = \beta_0 + \delta_1 d87_t + \delta_2 d92_t + \beta_1 tax_{it} + \cdots + a_i + u_{it}$$

式中，$d87_t$ 和 $d92_t$ 为年虚拟量；tax_{it} 为郡县 i 在时间 t 的税率（用百分数形式）。在方程中我们会有其他随时间而变的量，包括从事商业活动成本的度量（如平均工资）、工人生产能力的度量（如以平均教育度量）等等。a_i 是固定影响，包括所有不随时间而改变的因素，u_{it} 是特异误差项。为消除 a_i，我们可以对这些年的数据求差分，或者利用除掉均值后的时间序列（固定效应变换）。

数　据

通常你应该用一个章节来仔细描述你在经验估计中用到的数据。如果你的数据是非标准的或是未被其他研究人员广泛使用的，这就尤其必要了。原则上，你应该展示足够的信息以使读者可以获得数据并重新进行你的分析。特别是，所有适当的公共数据来源都应包括在参考书目中，而且小的数据集也应该列在附录中。如果你是通过自己调查搜集到的数据，在附录中也应给出一份问卷。

在讨论数据来源的时候，要明确每个变量的单位。（例如，收入是用百美元还是千美元计量的？）列出一张变量定义的表格对读者来说也非常有用。在表中出现的名字应与后面的章节中用以描述计量经济学结果的名字相一致。

一张总结统计量、如最小值、最大值、均值和每个变量的标准差的表格也是满含信息的。有这样一张表使得解释后面章节的系数估计值变得更简单，而且它强调了变量的测量单位。对二值变量来说，惟一需要的统计总结是样本中取值为 1 的比例（等同于样本均值）。对趋势变量来说，诸如均值

的统计量是不太令人感兴趣的。而对样本中的某一变量求出多年的平均增长率则常常会是有用的。

631 通常，你应该清楚地表明进行了多少次观测。对时间序列数据集来说，要明确你在分析中用到的年期，包括对历史上任何特殊时期（如第二次世界大战）的描述。如果你利用一个混合截面或一个综列数据集，一定要记录每年中有多少个横截面数据单元（人、城市等）。

结 果

结果章节应包括你在模型建立一节中的任何模型的估计。你可以从一个非常简单的分析开始。例如，假使毕业班学生进入大学的百分比（*percoll*）被用来度量某人就读的高中质量，那么，一个待估计的方程为

$$\log(wage) = \beta_0 + \beta_1 percoll + u$$

当然，它并未控制既决定工资又与 *percoll* 相关的一些其他因素。但是，一个简单的分析可以将读者带入一个更为深奥微妙的分析中并揭示对其他变量进行控制的重要性。

如果仅仅估计了几个方程，你可以用方程的形式把结果展示出来，标准误用圆括号括起来置于已估计出的系数之下。如果你的模型有几个解释变量并且你给出了一般模型的一些变异形式，那么用表格的形式记录结果就比方程的形式更好。大多数情形都至少应该有一张表，包括至少每个方程的 R-平方和观测次数。其他统计量，如校正的 R-平方，也可以列出来。

最重要的事情是讨论你对经验结果的解释及其力度。系数的符号是我们所期望的吗？它们统计上显著吗？如果一个系数统计上显著但有着与直觉相违的符号，这为什么可能是真的呢？它或许揭示了数据或计量经济学方法的一个问题（比方说，因为遗漏变量问题，OLS或许是不适宜的）。

务必描述主要解释变量系数的大小。通常有一个或两个政策变量对研究非常重要，它们的符号、大小和统计显著性都必须仔细对待。切记要区别开经济学和统计学上的显著性。如果一个 t 统计量很小，就应该弄清楚，是因为系数实际上很小还是因为它的标准误很大。

除了讨论一般模型中的估计值，你还可以提供一些有趣的特例，特别是那些需要检验某些多重假设的例子。例如，在确定产业间工资差异的研究中，你可以在没有产业虚拟变量的情况下给出方程；这可以使读者易于检验产业差异是否统计上显著（利用 F 检验的 R-平方形式）。不要过于担心为发现“最优”解释变量组合而舍弃某些变量的问题。正如我们早先提及的，这是一项困难而且甚至尚未较好定义的工作。只有在消除一组变量实质性地改变所关注系数的大小和/或显著性时，这一点才是重要的。舍弃一组变量来简化模型——如二次式或相互作用——可以通过 F 检验来证明。

如果你利用了至少两种不同的方法——如OLS和2SLS，或时间序列的

632 水平值和差分，或混合 OLS 和综列数据集的差分——那么你应该就任何关键性的差异作出评价。特别是，如果 OLS 给出了与直觉相违的结果，那么 2SLS 或综列数据方法能够改进估计值吗？

结 论

这可以是简短的一节以总结你所学到的东西。比方说，你也许想阐述你格外关注的系数的大小。结论中还应讨论所得结论会有什么蹊跷，以及建议进一步研究的可能方向。不妨设想读者要先看结论再决定是否阅读论文剩下的部分。

风格提示

你应该给你的论文一个足以反映其主题的名称。论文打印应留足行距。对所有的方程式都必须另起一行，居中并且连续编号，即编为（1）、（2）、（3）等等。大的图和表应放在正文后面。在文章中，用作者和时间进行索引，如 White（1980）。论文末的索引章节应采用标准格式。本书末的索引给出了一些例子。

当你在“计量经济学模型”一节中引入一个方程时，你应该说明重要的变量：因变量和一个或多个主要的自变量。如果你想着重于某一自变量，你可以如下写出一个方程

$$GPA = \beta_0 + \beta_1 alcohol + \boldsymbol{x\delta} + u$$

或

$$\log(wage) = \beta_0 + \beta_1 educ + \boldsymbol{x\delta} + u$$

记号 $\boldsymbol{x\delta}$ 是其他几个解释变量的缩写。在这一点上，你只需要一般地说明它们；在数据章节中可用一张表对它们进行明确说明。举例来说，在影响首席执行官收入的研究中，你可以在数据章节中包括下面这张表。

表 1	**变量说明**
salary	1990 年的年收入（包括津贴）（千美元）
sales	1990 年公司销售额（百万美元）
roe	1988—1990 年平均净资产回报率（百分数）
pcsal	1988—1990 年工资变化百分数
pcroe	1988—1990 年平均净资产回报率变化百分数

633

indust	如果是工业企业则为 1，否则为 0
finance	如果是金融企业则为 1，否则为 0
consprod	如果是消费产品企业则为 1，否则为 0
util	如果是公用事业企业则为 1，否则为 0
ceoten	担任公司 CEO 的年数

利用数据集 401K.RAW，来研究影响401k 养老金计划参与率的因素，有关的摘要统计量可建立如下。

表 2　　摘要统计量

变量	均值	标准差	最小值	最大值
prate	0.869	0.167	0.023	1
mrate	0.746	0.844	0.011	5
employ	4 621.01	16 299.64	53	443 040
age	13.14	9.63	4	76
sole	0.415	0.493	0	1
观测次数 = 3 784				

在结果一节中，你可以用方程的形式把估计值写出来，就像我们通常所做的那样。但也可以用表格展示；特别是当用不同的解释变量集估计出一个模型时，表格非常有用。如果你用方程写出估计结果，例如

$$
\begin{aligned}
\log(\widehat{salary}) &= \underset{(0.93)}{2.45} + \underset{(0.115)}{0.236}\log(sales) + \underset{(0.003)}{0.008}\,roe + \underset{(0.028)}{0.061}\,ceoten \\
n &= 204,\ R^2 = 0.351
\end{aligned}
$$

一定要在接近第一个方程的地方说明标准误已标明在括号里。为检验 $H_0: \beta_j = 0$，报告 t 统计量或它们的绝对值也是可以接受的，但最重要的是说明你在做什么。

如果你用表格形式给出结果，则要求因变量和自变量都被清楚地标明，634 并且把标准误或者 t 统计量（前者更好）注明在系数下方。有些作者喜欢用星号来表明不同显著水平下的统计显著性（例如，一颗星表示 5% 的显著水平，两颗星表示 1% 而不是 5% 的显著水平等等）。如果你在行文中仔细讨论了解释变量的显著性，这就不必要了。

报告结果的一个样本表格如下。

表 3 **OLS 结果**

因变量：参与率

自变量			
mrate	0.156	0.239	0.218
	(0.012)	(0.042)	(0.342)
$mrate^2$	—	-0.087	-0.096
		(0.043)	(0.073)
log (*emp*)	-0.112	-0.112	0.098
	(0.014)	(0.014)	(0.111)
$\log(emp)^2$	0.005 7	0.005 7	0.005 2
	(0.000 9)	(0.000 9)	(0.000 7)
age	0.006 0	0.005 9	0.005 0
	(0.001 0)	(0.001 0)	(0.002 1)
age^2	-0.000 07	-0.000 07	-0.000 06
	(0.000 02)	(0.000 02)	(0.000 02)
sloe	-0.000 1	0.000 8	0.000 6
	(0.005 8)	(0.005 8)	(0.006 1)
constant	1.213	0.198	0.085
	(0.051)	(0.052)	(0.041)
产业虚拟变量	否	否	是
观测次数	3 784	3 784	3 784
R-平方	0.143	0.152	0.152

说明：估计值下面用括号括起来的量是标准误。

635 如果你选择了因变量和自变量二者的单位以使系数不至于过大或者过小，你的结果就更易于阅读和解释。对于系数或标准误，永远也不要用诸如 1.051*e*-007 或者 3.524*e*+006 的数字报告出来，而且也不应该用科学符号。如果系数极其小或者极其大，就要像我们在第 6 章中讨论过的那样重新度量因变量或自变量。你还应该限制小数点后的位数。例如，如果你的回归软件包估计一个系数为 0.548 210 59，你就应该在论文中将其报告为 0.548 或者甚至是 0.55。

作为一般性规则，对用以产生结果的特定计量经济学软件包的命令，不应该在论文中出现；只有结果才是重要的。如果某一特殊命令是用以执行一

个特定的估计方法，可以在附录中将其介绍。附录也是放置支持你的分析但并非极其重要的附加结果的好地方。

▶ 小结

在这一章中，我们讨论了一个成功的经验研究的要素，还给出了能够提高分析质量的建议。而最终任何研究的成功，关键都要靠细心和投入的努力。

关键术语

数据开采
因特网
网上数据基础
网上搜寻服务
敏感度分析
总分析表
文本编辑器
文本（ASCII）文件

样本经验项目

通过本书的学习，我们已经看到了许多计量经济研究的实例，这些例子或者来自已发表的作品或者是由这些作品引发出来的作业。希望这些能使你对经验分析的范围有一个完整的概念。我们把以下内容作为额外的例子列出来，也许某些人会觉得它们值得关注。目的是激发你的想像力；我们不打算给出特定模型、数据要求或者可替换的估计方法的所有细节。用一个学期去完成这些项目是可能的。

636 1. 亲自进行一个校园调查来回答你所关注的你的大学里的问题。例如，工作对大学 *GPA* 的影响是怎样的？你可以就高中 *GPA*、大学 *GPA*、*ACT* 或者 *SAT* 分数、每周工作的小时数、体育运动的参与情况、主修的专业、性别和种族等问题对同学们提问。然后用这些变量来创建一个解释 *GPA* 的模型。每周增加一小时的工作时间对 *GPA* 产生多大的影响（如果有的话）？一个值得关注的问题是工作时间或许是内生的：它或许与影响大学 *GPA* 的不可观测因素相关，或者较低的 *GPA* 会导致学生做更多的工作。

更好的方法是选择这个学期之前的累积 *GPA*，然后得到最近这个学期的 *GPA*、这个学期的工作量以及其他变量。现在，在方程中使用累积 *GPA* 作为控制变量（解释变量）。

2. 在前面的题目中可以有许多变异形式。你可以研究药物或者酒精用量或者生活在社区中对平均积点分的影响。你也许要控制许多家庭背景变量及过去表现变量。

3. 城市的枪支控制法能够减少暴力犯罪吗？这样的问题很难用单一的截面来回答，因为城市和州的法律通常是内生的。[以克莱克和帕特森（Kleck and Patterson，1993）为例。他们利用了横截面数据和工具变量的方法，但他们的 IV 是值得质疑的。] 综列数据在推断这种背景之下的因果关系时非常有用。至少，你可以控制前一年的暴力犯罪率。

4. 洛和麦克菲特（Low and McPheters，1983）利用了工资率、警察死亡风险的估计值和其他一些控制变量的城市横截面数据。想法是，在有更高的因公受伤或死亡风险的城市工作，确定警察是否因此而得到了补偿。

5. 父母同意法会增加青少年的生育率吗？为此你可以利用州一级的数据：或者一个给定州的时间序列，或者更好地，各州的综列数据集。相同的法律降低了青少年的生育率吗？美国统计摘要包括所有州一级的数据。利瓦因、特雷纳和齐默尔曼（Levine，Trainor and Zimmerman，1996）研究了流产基金约束对类似结果的影响。其他因素，如流产的可能性，都可能影响青少年出生和流产率。

6. 交通法的改变会影响交通死亡率吗？麦卡锡（McCarthy，1994）含有对加利福尼亚州每月时间序列数据分析，可用一组虚拟变量来标示某些法律生效的月份。文件 TRAFFIC2.RAW 包括麦卡锡使用过的数据。一种替换方法是获取美国各州的综列数据集，由此你可以发现不同的州和不同时间的法律变化。(见文件 TRAFFIC1.RAW。)

马拉赫伊和辛德拉（Mullahy and Sindelar，1994）利用与州的法律相匹配的个人数据和酒税来估计法律和税收对醉酒驾车概率的影响。

7. 黑人在贷出市场上受到歧视吗？亨特和沃尔克（Hunter and Walker，1996）关注了这个问题；我们在习题 7.16 和 17.9 中利用了他们的数据。

8. 职业运动员有结婚奖励吗？考伦曼和努马克（Korenman and Neumark，1991）运用各种计量经济学方法发现了已婚男士显著的工资奖励。职业运动员——例如 NBA 运动员、重要的棒球手和职业的高尔夫球手——为
637 我们研究结婚奖励提供了一个有趣的群体，因为我们可以由此观察一些反映能力的度量。个人体育项目（如高尔夫和网球）的运动员，收入直接反映能力。在团队运动中，收入也许并不完全反映能力——比方说，在队里待的年数也会有影响。因此，在一个以诸如得分为因变量的方程中，我们可以加入一个结婚标示变量，而且在一个是因变量的回归中，一些能力控制量则属于自变量之列。

9. 回答问题：雪茄吸食者的生产能力较差吗？此问题的一个变形为：吸烟的工人会请更多的病假吗（其他情况均相同）？马拉赫伊和波特尼（Mullahy and Portney，1990）利用个人一级的数据评价了这一问题，而你可以利用诸如大城市一级的数据。诸如平均生产能力的因素可与吸烟工人的百分比相联系。其他变量，例如工人的平均教育水平、每个工人的资本和城市

规模（你还能想到更多）应为控制变量。

10．最低工资保障能缓解贫穷吗？你可以利用州或国家的数据来回答这个问题。观点是最低工资会随着州的不同而不同，因为有些州的最低工资比联邦政府的最低工资高。而且，一个州之内的名义最低工资会随时间而改变，有些是因为联邦政府工资水平的改变，而有些是因为州的工资水平的改变。努马克和韦斯切（Neumark ang Wascher，1995）利用州的综列数据估计了最低工资对年轻工人就业率的影响、还有对入学率的影响。

11．哪些因素会影响公立学校学生的表现？在大多数的州，要得到学校一级或至少区一级的数据相当简单。每个学生的花费会产生影响吗？师生比例会有影响吗？估计其他条件不变的情况下的影响较难，因为花费是与其他因素——例如家庭收入或贫穷比率——相关的。针对密歇根高中的数据集MEAP93.RAW含有对贫穷比率的度量。其他的可能方法是利用综列数据，或者至少控制前一年对学生表现的度量（如平均考试成绩或者通过了一门考试的学生的百分比）。

你可以考虑影响学生表现的不那么明显的因素。例如，在控制了收入之后，家庭结构会对这个问题产生影响吗？或许双亲中仅有一人有收入的家庭对学生表现有积极的影响。（至少有两种途径：家长花更多的时间与孩子待在一起，或者他们自愿到校上课。）在控制收入和其他因素的情况下，单亲家庭的影响是怎样的？你可以把一年或两年的普查数据与校区数据合并起来。

与更多的私立学校临近的公立学校会因为竞争而对学生的教育更好吗？这里有一个微妙的联立性问题，因为私立学校很可能坐落在公立学校已经较差的地区。豪克斯比（Hoxby，1994）利用了工具变量的方法，用不同宗教的人口比例作为私立学校数目的工具。

罗斯（Rouse，1998）曾研究一个不同的问题：根据米尔沃奇（Milwaukee）担保人方案，能够入读私立学校的学生是否比不能入读的学生表现更好？她利用了综列数据并得以控制一个不可观测的学生效应。

638 12．股票的超额回报或股票指数能够用滞后的价格股息比来预测吗？或者，用滞后利率或每周货币政策来预测？挑选一个国外股票指数或者是一个比较不著名的美国指数，这个分析将会非常有趣。为解释超额股票回报，科克伦（Cochrane，1997）作出了对新近的理论和经验结果的一个很好的调研。

13．在棒球卡市场上有种族歧视吗？这一分析将棒球卡与影响其价格的诸如职业统计、该球员是否杰出人物等等因素联系在一起。保持其他因素不变，黑人或西班牙裔运动员的卡是否被折价出售？

14．你可以检验体育运动的赌博市场是否有效。举例来说，足球或篮球的大幅广告包括了所有的有用信息，以至能够让人作出与广告相反的预测吗？数据集PNTSPRD.RAW包括了大学男子篮球比赛的信息。结果变量是二值的。即广告是否被包括进来了？于是，你可以在每场比赛之前试图获取已有的信息从而预测是否广告被包括进去了。（祝你好运！）

15. 大学体育运动的成功对学校的其他方面（申请、学生质量及非体育系）有何影响（如果有的话）？麦克密克和廷斯利（McCormick and Tinsley，1987）研究了一些主要学院的运动的成就对入学新生的 *SAT* 分数的影响。这里时间的选择很重要：不妨认为，刚结束的胜利影响了最近的申请和学生质量。我们必须控制其他许多因素——如学费和学校质量的度量——以使分析令人信服，因为如果不控制其他变量，在学术表现和运动成绩之间就会有负相关。

一个变异形式是为了配合足球或者男子篮球的自然竞争，而把赢得足球比赛或赢得一场或多场篮球比赛作为不同学校之间的差异的函数来看待。ATHLET1.RAW 和 ATHLET2.RAW 为可扩充和可刷新的两个小数据集。

16. 收集一个城市或者郡县样本某两年的谋杀率（比方说可以找 FBI 的统一犯罪记录）。选取经济和人口统计变量都易于从郡县或城市数据手册中获得的那一年作为后一年。从美国统计摘要中你可以得到死亡那一列的总人数，还有那一年的州一级的死刑量。如果年份是 1990 年和 1985 年，你可以估计

$$mrdrte_{90} = \beta_0 + \beta_1 mrdrte_{85} + \beta_2 executions + \text{其他因素}$$

其中我们关注的是 *executions* 的系数。滞后谋杀率和其他因素是控制变量。

其他因素也可以作为犯罪的制止因素。举例来说，克劳宁格（Cloninger，1991）给出了警方对犯罪率作出致命反击的效果的横截面分析。

639 作为一个类比，我们可以考虑，哪些因素影响大学校园的犯罪率？生活在兄弟会或者女学生联谊会中的学生的比例会对其产生影响吗？警方力量的大小或者采用的监管方式会起作用吗？（在这里推断因果关系的时候要小心。）采取护送计划有助于减少犯罪吗？邻近社区的犯罪率有影响吗？最近，各学院和大学被要求报告其犯罪统计；而在早些年，是否报告完全是自愿的。

17. 哪些因素会影响州一级的制造业生产率呢？除了资本和工人的教育水平，你还可以考虑一下工会化的程度。利用两个统计调查年（比方说 1980 年和 1990 年）的一个综列数据分析在这里或许是最具说服力的。克拉克（Clark，1984）给出了一个关于工会化程度如何影响公司业绩和生产率的分析。还有其他哪些变量可能解释生产率？

公司一级的数据可以从 *Compustat* 获得。例如，你可以研究在其他因素固定时工会化的改变是否会影响一个公司的股票价格。

18. 利用州一级、国家一级的数据，或者如果可能的话，利用校区一级的数据来研究一下影响平均学生教育支出的因素。一个值得关注的问题是，其他条件相等（例如收入和居民的教育水平都不变），一个有较高比例的年长者的地区会为学校花费较少的钱吗？普查数据与校区支出数据配合起来可获得一个庞大的横截面。美国教育部收集了这类数据。

19. 州的规章，如佩带摩托车安全帽法规，对摩托车驾驶的死亡率有何影响？或者，划船法案的区别——如最低驾船年龄——有助于解释划船事故

率吗？美国交通部收集了这类信息。可以与美国统计摘要的数据结合起来进行研究。综列数据分析在这里看来是有道理的。

20. 哪些因素会影响产出的增长？两个值得关注的因素是通货膨胀和投资［见 Blomström, Lipsey, and Zejan（1996）］。你可以利用一个值得你关注的国家的时间序列数据。或者，你也可以利用多个国家的截面数据，就像德朗和萨默斯（De Long and Summers, 1991）一样。弗里德曼和库特内（Friedman and Kuttner, 1992）发现有迹象表明，至少在 20 世纪 80 年代，商业票据利率与国库券利率之间的差异影响了实际产出。

21. 在美国经济（或其他经济）中，公司合并是怎样一种行为？通过展现两对数之差，大体上说就是增长率，沙哈特和托里森（Shughart and Tollison, 1984）把美国经济中每年的合并量（的对数）看做一个随机游走：在给定过去的增长率的情况下它（指合并量的对数）是不可预测的。这一结论仍然成立吗？它在不同的行业间均成立吗？过去对经济活动的何种度量可以用来预测合并活动？

22. 哪些因素可以解释就业和工资中的种族和性别差异？举例来说，霍尔泽（Holzer, 1991）重审了“空间不协调假说（spatial mismatch hypothesis）”来解释黑人和白人就业率上的差异。考伦曼和努马克（Korenman and Neumark, 1992）研究了分娩对妇女工资的影响，而赫斯奇和斯特拉顿（Hersch and Stratton, 1997）则关注家庭负担对男女工资的影响。

640 23. 收集青少年就业率、最低工资和影响青少年就业的因素的月份或季度数据来估计最低工资对青少年就业的影响。索伦（Solon, 1985）利用了美国的季度数据，而卡斯蒂罗-弗里曼和弗里曼（Castillo-Freeman and Freeman, 1992）利用了波多黎各的年数据。分析美国低收入州的时间序列数据或许是富于信息的——在美国，最低工资的变化将极有可能产生最大的影响。

24. 在城市的层面上估计犯罪率的时间序列模型。克劳宁格和萨特里斯（Cloninger and Sartorius, 1979）就是一例。作为一种新的窍门，你可以估计社区巡逻或者午夜篮球比赛——较新的反犯罪创举——的效果。因果关系的推断是微妙的。把一个滞后因变量包括进来或许有所帮助。因为你所使用的是时间序列数据，所以要当心谬误回归问题。

格罗戈尔（Grogger, 1990）利用每日杀人罪的数据来估计死刑的威慑效果。是否还有其他因素——比方说关于警方致命反击的新闻——也对每天的犯罪的案件数有影响？

25. 电脑的使用会产生总生产力效果吗？你需要取得关于生产力、使用电脑的雇员的百分比和其他因素的——或许是国家范围的——时间序列数据。用于研究和发展的花费（也许作为总销售额的一部分）是否应该考虑进来？有哪些社会因素可能影响生产力？饮酒量吗？还是离婚率？

26. 哪些因素会影响首席执行官的薪金？文件 CEOSAL1.RAW 和 CEOSAL2.RAW 含有用多种方式度量的公司业绩以及诸如任期和教育信息的数据集。你当然也可以更新这些数据文件或者寻找其他值得关注的因素。

罗斯和谢泼德（Rose and Shepard，1997）将公司多样化作为 CEO 报酬的一个重要的决定因素。

27. 各州之间税法的差异会影响国外直接投资的总量吗？海因斯（Hines，1996）研究了州的公司税以及申请国外税收贷放的能力对吸收美国之外的投资的影响。

28. 哪些因素影响选举结果？选举支出与其有关系吗？对特定问题的投票与其有关系吗？当地的经济状况呢？例如，参见莱维特（Levitt，1994）和数据集 VOTE1.RAW 和 VOTE2.RAW。弗尔（Fair，1996）做过美国总统大选的一个时间序列分析。

期刊列表

以下列举了人们熟知的部分期刊，它们包括关于经验企业、经济和其他社会科学的研究。所有这类期刊都可以在因特网上找到。

American Economic Review

American Journal of Agricultural Economics

American Political Science Review

Applied Economics

Brookings Papers on Economic Activity

Canadian Journal of Economics

Demography

641 *Economic Inquiry*

Economica

Economics Letters

Empirical Economics

Federal Reserve Bulletin

International Economic Review

Journal of Applied Econometrics

Journal of Business and Economic Statistics

Journal of Development Economics

Journal of Economic Education

Journal of Empirical Finance

Journal of Environmental Economics and Management

Journal of Finance

Journal of Health Economics

Journal of Human Resources

Journal of Industrial Economics

Journal of International Economics

Journal of Labor Economics
Journal of Political Economy
Journal of Public Economics
Journal of Monetary Economics
Journal of Money, Credit, and Banking
Journal of Quantitative Criminology
Journal of Urban Economics
National Bureau of Economic Research Working Paper Series
National Tax Journal
Public Finance Quarterly
Quarterly Journal of Economics
Regional Science & Urban Economics
Review of Economic Studies
Review of Economics and Statistics

数据资源

全世界有无数可用数据资源。大多数国家的政府都收集着大量的数据；在美国，普及且易于获得的数据资源，例如 *Economic Report of the President*，*the Statistical Abstract of the United States* 以及 *County and City Data Book*，我们都已经提及。许多国家的国际金融数据每年都会在 *International Financial Statistics* 上发表出来。各种杂志，像 *Business Week* 和 *U. S. News and World Report*，通常会发表统计数据——例如 CEO 的薪金和公司业绩，或者学术科目的排名——这些数据非常新而且能够在计量经济学分析中得到应用。

642 与试图提供列表的做法不同，我们列出一些因特网地址，为经济学家提供全面的资源。对经济学家来说，一个很有用的站点为 Resources for Economists on the Internet，它是由南密西西比大学的 Bill Goffe 维护的。地址是

http://econwpa.wustl.edu/EconFAQ/EconFAQ.html.

这个网站给出了期刊、数据资源以及专业和理论经济学家的列表的链接。使用起来非常简单。

美国统计协会的商业和经济统计部拥有极其详尽的数据资源列表并给出了它们的链接。地址是

http://www.econ-datalinks.org.

另外，*Journal of Applied Econometrics* 和 *Journal of Business and Economics Statistics* 还有一些数据档案，它们包括了近年来发表的大部分论文中用到的数据集。如果你找到了令你感兴趣的数据集，那么这是一条非常好的

可行之路，因为其中大部分数据的清理和格式化的工作已经完成了。不利之处在于，这些数据集的一部分是用于比我们在本书中所学更高深的计量经济学分析中的；另一方面，用标准计量经济学方法估计较简单的模型来作为比较，通常会有帮助。

许多大学，例如加利福尼亚大学伯克莱分校、密歇根大学和马里兰大学，都拥有大量的数据集以及对各种数据集的链接。你所在学校的图书馆说不定有大量关于商业、经济和其他社会科学的数据库的链接。地方联邦储备银行，例如圣·路易斯（St. Louis）的那个，掌握了各种数据。National Bureau of Economic Research 会把它的部分研究人员所用的数据集刊登出来。现在州和地方政府也发表大量的数据，而这些数据可以通过因特网找到。普查数据可以通过 Department of Census 公开得到。（两份有用的刊物分别是逢以 2 和 7 结尾的年数出刊的 *Census of Manufacturing*，和逢每十年的开始出刊的 *Census of the Population*。）其他机构，例如 *Department of Justice*，同样也为大众提供数据资源。

附录 A 基本数学工具

648 本附录包括一些计量经济学分析中用到的基本数学,我们扼要论述了总和运算子的各种性质,探讨了线性和某些非线性方程的性质,并复习了比例和百分数。我们还介绍了一些在应用计量经济学中常见的特殊函数,包括二次函数和自然对数,前4节仅涉及基本的代数技巧,A.5节则是微分学的一个简要回顾;虽然要读懂本课文的大部分,微积分并不是必需的,但在章末附录和第Ⅲ部分的一些较高深的章节中我们用到了微积分。

A.1 总和运算子与描述统计量

总和运算子(summation operator)是用以表示多个数的总和运算的一个有用的缩写符号,它在统计学和计量经济学分析中扮演着重要角色。如果 $\{x_i: i = 1,\cdots,n\}$ 表示 n 个数的一个序列,那么我们就把这 n 个数的总和写为

$$\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \cdots + x_n \tag{A.1}$$

按照这个定义,容易表明总和运算子有下列性质:

性质 SUM.1

对任意常数 c，有

$$\sum_{i=1}^{n} c = nc \tag{A.2}$$

性质 SUM.2

对任意常数 c，有

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i \tag{A.3}$$

性质 SUM.3

如果 $\{(x_i, y_i): i = 1, 2, \cdots, n\}$ 是 n 双数的一个集合，并且 a 和 b 是常数，那么

$$\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i \tag{A.4}$$

644 还要注意，有些情况是不能通过总和运算子得到的，令 $\{(x_i, y_i): i = 1, 2, \cdots, n\}$ 为 n 双数的一个集合，且对每一 i，$y_i \neq 0$，那么

$$\sum_{i=1}^{n}(x_i/y_i) \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

换言之，比率的总和不等于总和的比率，当 $n = 2$ 时，利用熟知的初等代数知识也知道这个等式是不成立的：$x_1/y_1 + x_2/y_2 \neq (x_1 + x_2)/(y_1 + y_2)$。类似地，平方的总和不等于总和的平方：$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$，除非是特殊情况。取 $n = 2$ 时最容易看出这两个量一般地说是不相等的：$x_1^2 + x_2^2 \neq (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$。

给定 n 个数 $\{x_i: i = 1, \cdots, n\}$，把它们加起来再除以 x，便算出它们的平均数或均值：

$$x = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{A.5}$$

当这些 x_i 是某一变量（如受教育年数）的数据的一个样本时，我们常把这个均值叫做样本均值，以强调它是从数据的一个特殊集合计算出来的。样本均值是描述统计量的一个例子：这个统计量描述了点 x_i 的集合趋势。

关于均值，有些有必要弄明白的基本性质。首先，假定我们取 x 的每一观测值并从中减去其均值：$d_i \equiv x_i - \bar{x}$（这里“d”表示对均值的离差）。那么，这些离差之和必为零：

$$\begin{aligned}\sum_{i=1}^{n} d_i &= \sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \\ &= \sum_{i=1}^{n} x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0\end{aligned}$$

或扼要地写为

$$\sum_{i=1}^{n}(x_i-\bar{x})=0 \tag{A.6}$$

可用一个简单的例子说明这是怎么回事。假定且 $n=5$ 且 $x_1=6$，$x_2=1$，$x_3=-2$，$x_4=0$ 和 $x_5=5$，那么 $\bar{x}=2$，并且去掉均值的样本是 $\{4,-1,-4,-2,3\}$。这些值加起来便等于零，正如方程 A.6 所说的。

645 在第 2 章对回归分析的论述中，我们还需要知道有关离差的更多的代数运算结果，一个重要的事实是，离差平方和等于 x_i 的平方和减去 $\bar{x}$ 平方的 n 倍：

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n(\bar{x})^2 \tag{A.7}$$

利用总和运算子的基本性质就能证明这一点：

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2&=\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)\\&=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n(\bar{x})^2\\&=\sum_{i=1}^{n}x_i^2-2n(\bar{x})^2+n(\bar{x})^2\\&=\sum_{i=1}^{n}x_i^2-n(\bar{x})^2\end{aligned}$$

给定两个变量的数据集 $\{(x_i,y_i):i=1,2,\cdots,n\}$，可以证明：

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})&=\sum_{i=1}^{n}x_i(y_i-y)\\&=\sum_{i=1}^{n}(x_i-x)y_i=\sum_{i=1}^{n}x_iy_i-n(\bar{x}\cdot\bar{y})\end{aligned} \tag{A.8}$$

这是对方程(A.7)的一个推广[(在 A.7)中，对所有 i 都有 $y_i=x_i$]。

在本书的大部分中，我们都把均值作为集中趋势的主要度量看待，但有时用中位数或**样本中位数**(sample median)表示中心值也是有信息性的。为了得到 n 个值($x_1,\cdots,x_n$)的中位数，我们先把 x_i 值按从小到大的顺序排列。然后，如果 n 是奇数，样本中位数就是按顺序居中的那个数。例如，给定的一些数$\{-4,8,2,0,21,-10,18\}$，中位数就是 2(因为大小顺序的序列是$\{-10,4,0,2,8,18,21\}$)。如果我们把所列的最大值 21 改成它的 2 倍即 42，中位数仍然是 2，作为对比，样本均值则要从 5 变到 8，是一个不小的变化。一般地说，中位数和均值相比不那么受数列中极端值(大的或小的)变化的影响。这就是为什么在摘要报导某城市或某地县的收入或房价时，常常报告"中位数收入"或"中位数房价"，而不报告其均值的原因。

如果 n 是偶数，那么居中的就有两个数，将没有定义中位数据的惟一方法。通常把中位数定义为两个居中数的均值(仍指从小到大排序的数列)。利用这一规则，数集$\{4,12,2,6\}$的中位数就是$(4-6)/2=5$。

A.2 线性函数的性质

线性函数由于容易解释和运算而在计量经济学中扮演重要角色，如果两个变量和的关系是

646

$$y = \beta_0 + \beta_1 x \tag{A.9}$$

就说 y 是 x 的线性函数，而 β_0 和 β_1 是描述这个关系式的两个参量（常数），β_0 为截距，β_1 为斜率。

一个线性函数的定义特色在于，y 的改变量总是 x 的改变量的 β_1 倍：

$$\Delta y = \beta_1 \Delta x \tag{A.10}$$

式中，Δ 表示“改变量”。换句话说，x 对 y 的边际效应是一个等于 β_1 的常数。

例 A.1 线性住房支出函数

假定每月住房支出和每月收入的关系式是

$$housing = 164 + 0.27 income \tag{A.11}$$

那么，每增加 1 美元收入，就有 27 美分用于住房开支，如果家庭收入增加 200 美元，那么住房开支就增加(0.27)200 = 54 美元。图 A.1 描绘了这个函数的图形。

按照方程（A.11)，一个没有收入的家庭按字面解释，也有 164 美元的住房开支，这当然是不真实的，对低收入水平，这个线性函数不会很好地描述住房和收入之间的关系，这就是为什么我们最终还得用其他类型的函数来描述这种关系。

在方程（A.11）中，把收入用于住房的边际消费倾向（MPC）是 0.27。它不同于平均消费倾向（APC)，后者是

$$\frac{housing}{income} = \frac{164}{income} + 2.7$$

APC 并非常量，它总比 MPC 大，但随着收入的增加而接近 MPC。

多于两个变量的线性函数也是容易定义的，假定 y 与两个变量 x_1 和 x_2 有一般形式的关系：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{A.12}$$

这个函数的图形是三维的，较难以想像，不过 β_0 仍然是截距（即 $x_1 = 0$ 和 $x_2 = 0$ 时 y 的取值)，并且 β_1 和 β_2 都是具体斜率的度量。由式（A.12)，对给定的 x_1 和 x_2 的改变值，y 的改变值是

$$\Delta y = \beta_1 \Delta x_1 + \beta_2 \Delta x_2 \tag{A.13}$$

647

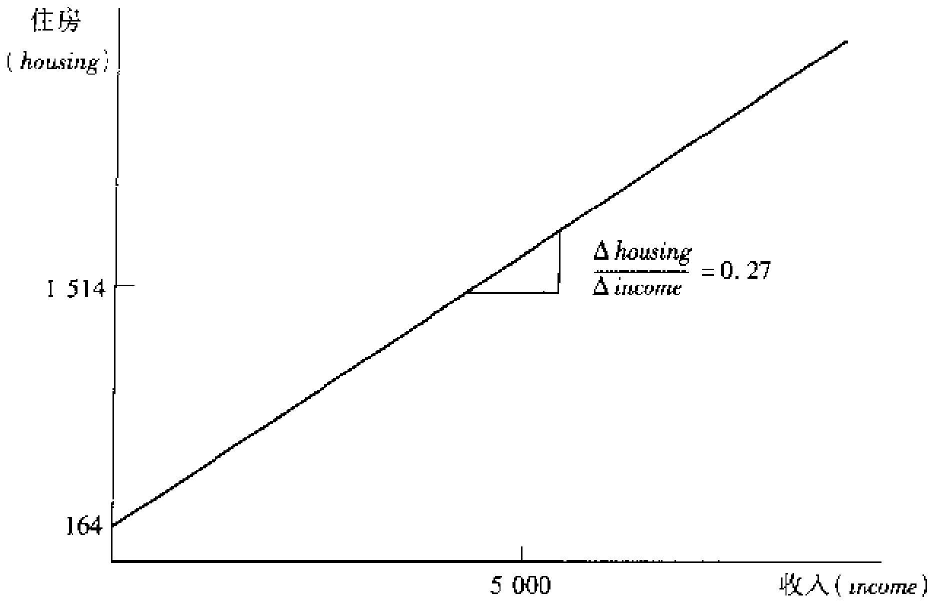


图 A.1　housing = 164 + 0.27income 的图形

如果 x_2 不改变，就是，说 $\Delta x_2 = 0$，则有

$$\Delta y = \beta_1 \Delta x_1 \text{，如果 } \Delta x_2 = 0$$

因此 β_1 是关系式在 x_1 方向的斜率：

$$\beta_1 = \frac{\Delta y}{\Delta x_1} \text{，如果 } \Delta x_2 = 0$$

因为它度量着保持 x_2 固定时 y 怎样地随 x_1 而变，所以常把 β_1 叫做 x_1 对 y 的**局部影响（或偏效应）**（partial effect）。由于局部效应（或影响）涉及保持其他因素固定不变，所以它紧密地联系着“其余情况不变（ceteris paribus）”的理念，参量 β_2 可作类似的解释，即 $\beta_2 = \Delta y / \Delta x_2 (\Delta x_1 = 0)$，因此 β_2 是 x_2 对 y 的局部效应。

例 A.2　对密纹音碟的需求

假定大学生对密纹音碟的每月需求量与密纹音碟的价格和每月自行支配的收入有如下关系：

$$quantity = 120 - 9.8 price + 0.03 income$$

648 式中，*price* 指每碟的美元数；*income* 以美元计。需求由线表示在 *income*(和其他因素)保持不变的情况下 *quantity* 和 *price* 的关系。图 A.2 描绘它在收入水平为 900 美元时的二维图形。需求曲线的斜率 −9.8 是价格对数量的局部影响：保持收入固定不变，如果密纹音碟的价格增加 1 美元，那么需求量就下跌 9.8(我们把音碟只能按离散单位购买的事实抽样化或连续化了)。收入的增加只是使需求线向上迁移(改变了截距)，但斜率仍然一样。

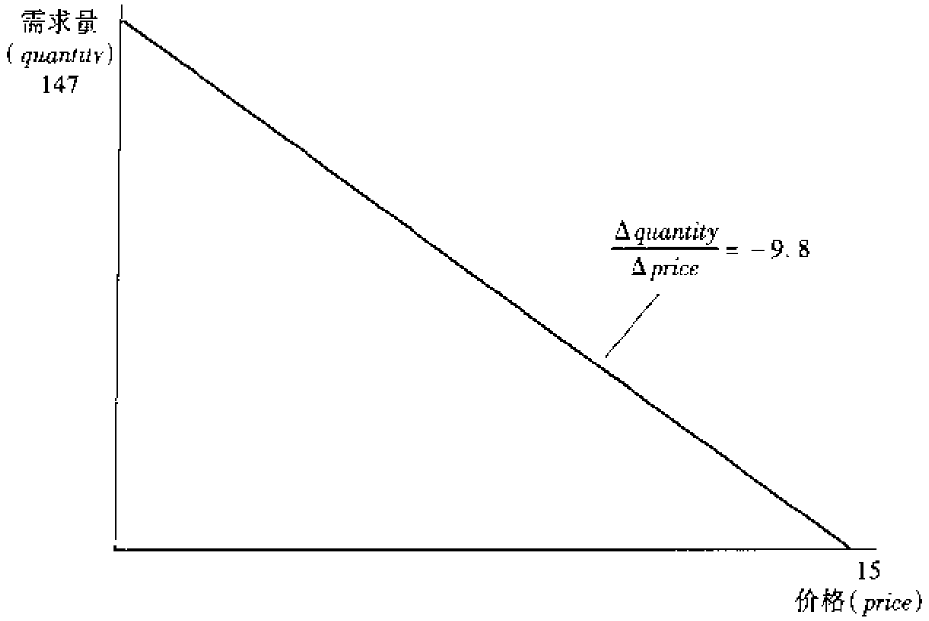


图 A.2 $quantity = 120 - 9.8price + 0.3income$,而且收入固定在 900 美元的图形

A.3 比例与百分数

比例与百分数在应用经济学中扮演如此重要的角色,以致我们有必要在工作中熟悉它们。在日常的报章中有许多的数量均以百分数的形式出现,如利息率、失业率和中学毕业生率就是其中的一些例子。

649 一种重要的技能就是知道怎样在比例和百分数之间进行转换。用 100 去乘一个比例就容易得到一个百分数,例如在一个县里有中学毕业程度的成年人占的比例是 0.82,那么就说 82% 的成年人有中学毕业学位。思考百分数和比例的另一方法是,比例是百分数的十进位形式。例如,年薪30 000美元的一个家庭的边际税率据报导为 28%,那么,再多 1 个美元收入的应付所得税比例就是 0.28(或 28 美分)。

在使用百分数时,我们常常需要把它转换成十进位形式。例如,一个州的销售税是 6%,而可课税项目的支出是 200 美元,那么,须付销售税是 200(0.06) = 12 美元。又如定期存款的年回报是 7.6%,而在年初你的存款是3 000美元,那么你的利息收入将是3 000(0.076) = 228 美元,而不是3 000乘以 7.6,尽管后者是你所希望的。

有时在大众媒体中误把比例报导为百分数,我们务必提防。如果看到"饮酒的中学生的百分数是 0.57",我们就知道这实际上是 57%(而不是像字面所表示的那样,仅仅比半个 1% 多一点)。大学排球迷也许熟悉报章的习惯用语,知道"她的击球成功百分数是 0.372"实际上是说她击球的成功百分数是 37.2%。

在计量经济学中,我们常对测量各种数量的变化感兴趣。令 x 代表某个变量,比如说个人收入、某社区中的犯罪次数或某厂商的利润。令 x_0 和 x_1 为 x 的两个值:x_0 是初始值,x_1 是随后的值。例如,x_0 是某人在 1994 年的年收入,而 x_1 是同一人在 1995 年的收入。那么 x 从 x_0 变到 x_1 的**比例变化**(proportional change)就是

$$(x_1 - x_0)/x_0 = \Delta x/x_0 \tag{A.14}$$

这里自然假定了 $x_0 \neq 0$。换句话说,为了得到比例变化,只须用 x 的初始值去除它的变化。这是把变化标准化为与单位无关的一种方法。例如一个人的收入从每年 30 000 美元变到每年 36 000 美元,则其比例变化是 6 000/30 000 = 0.20。

更常见的是用百分数表示变化,x 从 x_0 变到 x_1 的**百分数(或百分比)变化**(percentage change)就是

$$\%\Delta x = 100(\Delta x/x_0) \tag{A.15}$$

式中,记号 $\%\Delta x$ 读作"x 的百分数变化"。例如,当收入从 30 000 美元变到 33 750美元时,收入便增加了 12.5%;为了得到这个结果,只须用 100 去乘比例变化 0.125。

我们必须当心不要把百分数变化说成比例变化。比方说,在前面的例子中,把收入的百分数变化说为 0.125 是不正确的,这样会引起混乱。

当我们所考虑的变化是关于美元数或人口一类的变量时,百分数变化指的是什么并无含糊之处。相比之下,当变量本身就是百分数时,这是经济学和其他社会科学中常有的事情,再计算百分比变化就可能有些复杂。为了说明,令 x 代表某城市中受过大学教育的成年人的百分数。假定初始值是 $x_0 = 24$ 650 (24%受过大学教育),而新值是 $x_1 = 30$。利用这两个 x 值就能计算受大学教育人口百分数的变化。首先计算 x 的变化 Δx。本例中 $\Delta x = x_1 - x_0 = 6$:受大学教育人口的百分数已增加 6 个百分点。另一方面,我们可以利用方程(A.15)来计算 x 的百分数变化:$\%\Delta x = 100[(30-24)/24] = 25$。

在这个例子中,百分点变化和百分数变化很不一样。**百分点变化**(percentage point change)指这两个百分数的变化,而百分数变化则指相对于初始值的这一变化。总之,我们要十分注意所算的是哪一种变化。细心的研究者会把二者区分得清清楚楚;不幸的是,在学术研究中也和大众报章中那样,所说的变化是哪一类型往往含糊不清。

例 A.3　密歇根州提高销售税

1994 年 3 月,密歇根州选民批准销售税从 4%提高到 6%。在政治宣传中,该措施的支持者宣称,销售税增加了 2 个百分点,或每个美元增加了 2 美分。反对增加销售税的人则称销售税提高了 50%,两方面的说法都是正确的,只是测量销售税增加的方法各异。自然,每方都以有利于自身立场的方式宣传这一措施。

对于像薪金一类的变量,谈什么"薪金的百分点"则是无意义的,因为薪金不是以百分数衡量的。我们可以用美元或百分数来描述薪金的变化。

A.4 若干特殊函数及其性质

在 A.2 节中，我们复习了线性函数的基本性质。我们曾经指出，像 $y = \beta_0 + \beta_1$ 这样的线性函数有一个重要性质：不管 x 的起始值是什么，x 每变化一个单位都导致 y 的同样变化。如我们前面所注意到的，这等于说 x 对 y 的边际效应是常数，这对许多经济关系来说多少有点不真实。例如，边际报酬递减这个重要的经济概念就不符合线性关系式。

为了建立各种经济现象的模型，我们需要研究一些非线性函数。**非线性函数**（nonlinear function）的特点是，给定 x 的变化，y 的变化依赖于 x 的起始值。某些非性线函数常出现于经验经济学中，我们有必要知道怎样去解释
651 它们，对非线性函数的完整的理解将把我们引到微积分的领域中。这里我们仅对这种函数的最重要方面摘要而谈，一些推导细节留到 A.5 节再处理。

二次函数

刻画报酬递减率的一个简单方法是在线性关系式中加一个二次项。考虑方程式

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{A.16}$$

式中，β_0，β_1 和 β_2 为参量。当 $\beta_1 > 0$ 而 $\beta_2 < 0$ 时，y 和 x 之间的关系呈抛物线形，如图 A.3 所示，其中取 $\beta_0 = 6$，$\beta_1 = 8$ 和 $\beta_2 = -2$。

当 $\beta_1 > 0$ 且 $\beta_2 < 0$ 时，可以证明（利用下节的微积分学）函数的最大值出现于点

$$x^* = \beta_1 / (-2\beta_2) \tag{A.17}$$

例如，若 $y = 6 + 8x - 2x$（从而 $\beta_1 = 8$，$\beta_2 = -2$），则 y 的最大值出现于 $x^* = 8/4 = 2$ 处，并且这个最大值是 $6 + 8(2) - 2(2)^2 = 14$（见图 A.3）。

652 方程（A.16）意味着 x 对 y 有一个递减的**边际效应**（diminishing marginal effect），这是容易从它的图形看出的。假定我们从 x 的低值开始，然后把 x 增大一个 c（比方说）值。这和从 x 的较高值开始同样增大 c 相比，对 y 有较大的效应。事实上，一旦 $x > x^*$，x 再增加将使 y 减小。

说 x 对 y 有一个递减边际效应，无异于说在图 A.3 中函数的斜率随 x 增加而减小。虽然这能从图形看得清楚，但是人们常常想把斜率下降的速度加以量化。应用微积分知识便给出这个二次函数的近似斜率：对于 x 的"微小"变化，有

$$\text{斜率} = \frac{\Delta y}{\Delta x} \approx \beta_1 + 2\beta_2 x \tag{A.18}$$

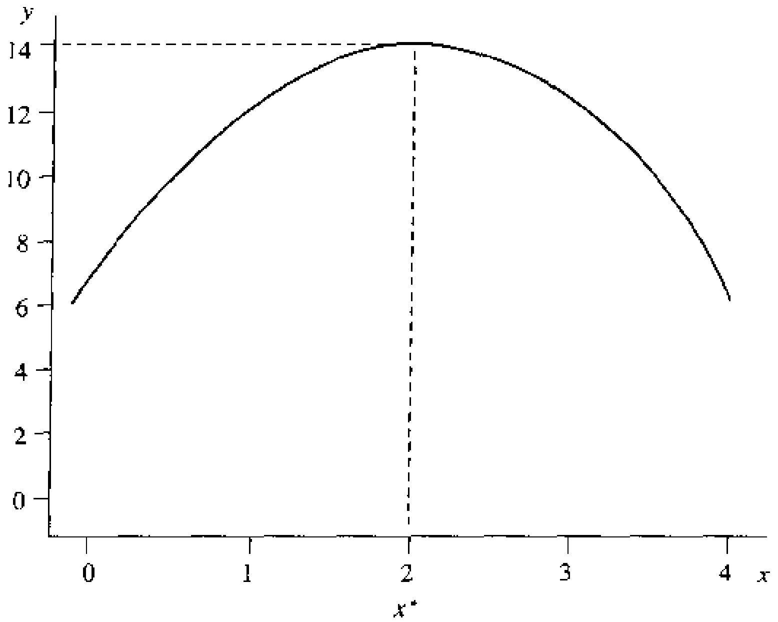


图 A.3　$y=6+8-2x^2$ 的图形

[方程（A.18）的右端是方程（A.16）中的函数对 x 的导函数（derivative）。] 写出这个关系式的另一方法是

$$\Delta y \approx (\beta_1 + 2\beta_2 x)\Delta x\text{，对于“微小的”}\Delta x \qquad \text{(A.19)}$$

为了看出这个近似到底有多好，再次以函数 $y=6+8x-2x^2$ 为例。根据方程(A.19)，$\Delta y\approx(8-4x)\Delta x$。现在，假定我们从 $x=1$ 开始并将 x 改变 $\Delta x=0.1$，利用式(A.19)，$\Delta y\approx(8-4)(0.1)=0.4$。当然，我们能够通过求 $x=1$ 和 $x=1.1$ 时的 y 值来准确地计算 y 的变化：$y_0=6+8(1)-2(1)^2=12$ 和 $y_1=6+8(1.1)-2(1.1)^2=12.38$，从而 y 的准确的变化是0.38。可见近似计算的 $\Delta y\approx0.4$ 是相当接近的。

现在，假定我们从 $x=1$ 开始但 x 改变一个较大的数：$\Delta x=0.5$，那么近似计算给出 $\Delta y\approx4(0.5)=2$，准确的变化可通过求 $x=1$ 和 $x=1.5$ 时的 y 来定出。前者的 y 值是12，而后者则是 $6+8(1.5)-2(1.5)^2=13.5$，从而实际变化是1.5（不是2）。由于 x 的变化较大，所以近似程度要差些。

在许多应用中，方程（A.19）都可用于计算取任何初始值和仅有微小变化的 x 对 y 的近似边际效应。

例 A.4　一个二次工资函数

假定每小时工资（*wage*）和参加劳动市场年数（代表工作经验 *exper*）之间的关系是

$$wage = 5.25 + 0.48exper - 0.008exper^2 \qquad \text{(A.20)}$$

此函数和图 A.3 中的函数有相同的一般形状。利用方程（A.17），经验 *exper* 一直到它的转折点 $exper^* = 0.48/[2(0.008)] = 30$ 为止，对工资都有正的

效应。工作的第一年经验大约值为 0.48 或 48 美分［用 $x=0$ 和 $\Delta x=1$ 代入 (A.19) 的结果］。每增加工作经验一年所增加的工资，都比先前一年所增加的要少——这反映了经验的递减边际报酬。在 30 年处再增加一年经验实际上将使工资降低。这并非很现实，但它是用二次函数刻画递减边际效应的 653 后果之一：在某点处，函数必然达到最大然后向下弯曲。为了实际的目的，最大值常出现在一个足够大的点处，以致造成并不重要的后果，但也不总是这样。

如果 $\beta_1<0$ 且 $\beta_2>0$，式（A.16）的二次函数图形就呈现 U 的形状，这时就有一个递增的边际报酬，函数的最小值出现在点 $-\beta_1/(2\beta_2)$ 处。

自然对数

在计量经济分析中起着最重要作用的非线性函数是**自然对数**（natural logarithm）。本书中我们把自然对数记为

$$y=\log(x) \tag{A.21}$$

这也就是我们经常说的**对数函数**（log function），你也许记得用过不同符号表示这个自然对数；最常用的是 $\ln(x)$ 或 $\log_e(x)$。当对数使用着几个不同的底数时，这些不同的符号是有用的。为了我们的目的，只有自然对数是重要的，因此我们用 $\log(x)$ 表示自然对数贯穿全书。这和许多统计包的使用符号相符，虽然有些使用 $\ln(x)$［而且大多计算器使用 $\ln(x)$］。经济学家兼用 $\log(x)$ 和 $\ln(x)$，当你阅读应用经济学论文时知道这点是有用的。

函数 $y=\log(x)$ 仅对 $x>0$ 有定义，其图形见图 A.4。要知道 $\log(x)$ 的值是怎样得来的并不十分重要。为了我们的目的，不妨把这个函数看做一个黑箱子：我们可以将任何代入 $x>0$，即可从计算器或计算机获得 $\log(x)$。

有几件事从图 A.4 看是显然的，首先，当 $y=\log(x)$ 时，y 和 x 的关系展现一种递减边际报酬。对数和图 A.3 的二次函数的一个重要差别是，当 $y=\log(x)$ 时，x 对 y 永远没有负效应：函数的斜率随着 x 的增大越来越接近零，然而这个斜率永远到不了零，所以更不会是负的。

从图 A.4 看以下情况也是明显的：

$$\log(x)<0,\ 0<x<1$$
$$\log(1)=0$$
$$\log(x)>0,\ x>1$$

特别是，$\log(x)$ 可正可负。关于对数函数的一些有用的代数事实是

$$\log(x_1\cdot x_2)=\log(x_1)+\log(x_2),\ x_1,x_2>0$$
$$\log(x_1/x_2)=\log(x_1)-\log(x_2),\ x_1,x_2>0$$
$$\log(x^c)=c\log(x),\ x>0,\ c\text{ 为任意数}$$

654

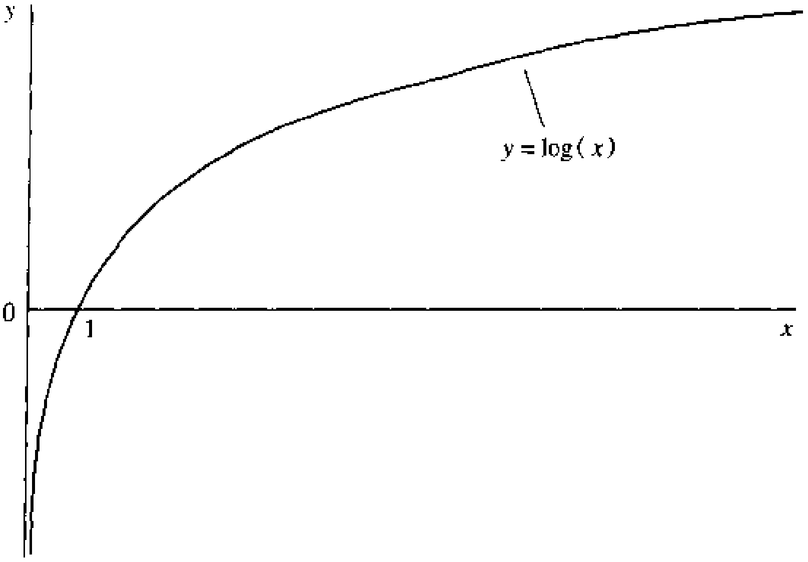


图 A.4　y = log(x)的图形

有时，我们必须依靠这些性质。

对数可用于计量经济学应用中的各种近似计算。首先，$\log(1+x) \approx x\ (x \approx 0)$，你不妨用 $x = 0.02, 0.1$ 和 0.5 试试看这个近似计算的质量随着 x 变大而恶化。更为有用的一个事实是，两对数之差可用做比例变化的近似值。令 x_0 和 x_1 为两个正数，可以证明（利用微积分学），对 x 的微小变化，有

$$\log(x_1) - \log(x_0) \approx (x_1 - x_0)/x_0 = \Delta x/x_0 \tag{A.22}$$

如果我们用 100 去乘方程(A.22)并写 $\Delta\log(x) = \log(x_1) - \log(x_0)$，那么，对 x 的微小变化，有

$$100 \cdot \Delta\log(x) \approx \%\Delta x \tag{A.23}$$

微小的含义有赖于具体情况。我们将会看到本书中的一些例子。

如果准确的百分数变化是如此容易计算的话，那么为什么要用式（A.23）去做近似计算呢？我们马上就能看到为什么式（A.23）的近似计算在计量经济学中是有用的。首先，我们用两个例子说明这种近似计算究竟有多好。 655

假定 $x_0 = 40$ 和 $x_1 = 41$，那么，利用$100(x_1 - x_0)/x_0$，x 从 x_0 变到 x_1 的百分数变化就是 2.5%。现在 $\log(41) - \log(40) = 0.024\ 7$(取 4 位小数)，若乘以 100 就非常接近 2.5。所以这个近似计算相当好。再考虑一个大得多的变化：$x_0 = 40$ 和 $x_1 = 60$，准确百分数变化是 50%，而 $\log(60) - \log(40) = 0.405\ 5$，从而近似值为 40.55%，这就远离目标了。

既然式（A.23）只适合于微小变化，它所给出的近似计算又为什么有用呢？为了作出有力的回答，我们首先定义 y 对 x 的**弹性**（elasticity）为

$$\frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{\%\Delta y}{\%\Delta x} \tag{A.24}$$

换句话说，y 对 x 的弹性就是当 x 增加 1% 时 y 的百分数变化。这个概念应从经济学入门课起就开始熟悉了。

如果 y 是 x 的线性函数：$y=\beta_0+\beta_1 x$，则这个弹性是

$$\frac{\Delta y}{\Delta x}\cdot\frac{x}{y}=\beta_1\cdot\frac{x}{y}=\beta_1\cdot\frac{x}{\beta_0+\beta_1 x} \tag{A.25}$$

这明显取决于 x 值。（这是基本需求理论的熟知结果的一个推广，弹性并不是沿着需求直线一路不变的。）

弹性不仅对需求理论，也对应用经济学许多领域是异常重要的。考虑一个**恒常**（constant）弹性模型，对许多研究来说都是方便的，而对数函数能帮助我们去设定这样的模型。如果我们把式（A.23）中的近似兼用于 x 和 y，那么这个弹性就近似地等于 $\Delta\log(y)/\Delta\log(x)$。因此，一个恒常弹性模型可近似地描述为方程

$$\log(y)=\beta_0+\beta_1\log(x) \tag{A.26}$$

式中，β_1 为 y 对 x 的弹性（假定 x，$y>0$）。

例 A.5　常弹性需求函数

如果 q 代表需求量而 p 代表价格，并且二者的关系为

$$\log(q)=4.7-1.25\log(p)$$

则需求的价格弹性是 -1.25。粗略地说，价格每增加 1%，将导致需求量的 1.25% 的下降。

---

656 为了我们的目的，式（A.26）中的 β_1 仅仅是接近于弹性这一事实并不重要。事实上，当我们用微积分学来定义弹性时——见 A.5 节——定义是精确的。为了计量经济分析的目的，式（A.26）定义了一个**恒常弹性模型**（constant elasticity model），这类模型在经验经济学中扮演着重要角色。

在经验研究工作中还常有用到对数函数的其他场合。假定 $y>0$ 并且

$$\log(y)=\beta_0+\beta_1 x \tag{A.27}$$

那么 $\Delta\log(y)=\beta_1\Delta x$，从而 $100\Delta\log(y)=(100\cdot\beta_1)\Delta x$，由此知，当 y 和 x 有方程（A.27）所表示的关系时，有

$$\%\Delta y\approx(100\cdot\beta_1)\Delta x \tag{A.28}$$

例 A.6　对数工资方程

假使每小时工资与受教育年数有如下关系：

$$\log(wage)=2.78+0.094\,educ$$

应用式(A.28)就有

$$\%\Delta wage\approx100(0.094)\Delta educ=9.4\Delta educ$$

由此知多受一年教育将使每小时工资增加约 9.4%。

通常把量$\%\Delta y/\Delta x$叫做y对x的**半弹性**(semi-elasticity),半弹性表示当x增加一单位时y的百分数变化。我们刚才所要表明的是,在模型(A.27)中,半弹性是个常数并且等于$100\cdot\beta_1$,在例A.6中,我们可以方便地把工资和教育的关系概括为:多受一年教育——不管从多少年教育算起——将使工资提高约9.4%。这说明这类模型为什么在经济学中起重要作用。

另一种关系式在应用经济学中也是有意义的:

$$y=\beta_0+\beta_1\log(x) \tag{A.29}$$

式中,$x>0$。怎样解释这个方程呢?如果作y的变化,将有$\Delta y=\beta_1\Delta\log(x)$,这又可写为$\Delta y=(\beta_1/100)[100\cdot\Delta\log(x)]$。这样,利用式(A.23)中的近似计算,我们就有

$$\Delta y\approx(\beta_1/100)(\%\Delta x) \tag{A.30}$$

换句话说,当x增加1%时y有$\beta_1/100$的单位变化。

657 **例A.7 劳动力供给函数**

假定一个工人的劳动力供给可描述为

$$hours=33+45.1\log(wage)$$

式中,$wage$为每小时工资;$hours$为每周工作小时数。于是,由式(A.30),有

$$\Delta hours\approx(45.1/100)(\%\Delta wage)=0.451\%\Delta wage$$

换言之,工资每增加1%将使每周工作小时增加约0.45,或者说略小于半个小时。如果工资增加10%,那么$\Delta hours=0.451(10)=4.51$,或者说约四个半小时。不宜对更大的工资百分数变化应用这个近似计算。

指数函数

在结束本节之前,仍有必要再讨论一种特殊函数,是与对数有关的一种。作为引子,且考虑方程(A.27)。那里,$\log(y)$是x的线性函数,但是怎样写出y本身作为x的一个函数呢?**指数函数**(exponential function)便给出这个答案。

我们把这个指数函数写为$y=\exp(x)$,它的图形由图A.5给出。

658 我们从图A.5看到,$\exp(x)$对任何x值都有定义,而且总是大于零的。有时人们把它写成$y=e^x$,但是在本书中我们不采用这个符号。指数函数的两个重要的数值是$\exp(0)=1$和$\exp(1)=2.7183$(取4位小数)。

指数函数在下述意义上是对数函数的逆函数:对所有的x有$\log[\exp(x)]=x$,而仅对$x>0$有$\exp[\log(x)]=x$。换言之,对数“解除了”指数,

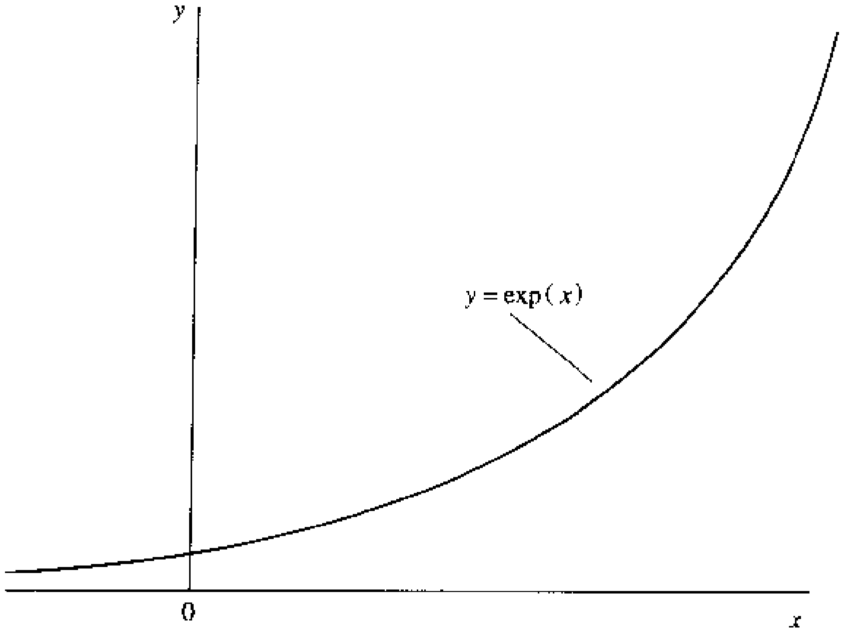


图 A.5　$y=\exp(x)$ 的图形

并且反之亦然。(这就是人们有时把指数函数叫做反对数函数的原因。) 特别地，要看到 $\log(y)=\beta_0+\beta_1 x$ 等价于

$$y=\exp(\beta_0+\beta_1 x)$$

如果 $\beta_1>0$，则 x 和 y 之间的关系就有如同图 A.5 一样的形状。因此，如果 $\log(y)=\beta_0+\beta_1 x$ 且 $\beta_1>0$，则 x 对 y 有一递增的边际效应。对于例 A.6，这就是说，再受一年教育要比前面一年受的教育导致更大的工资变化。

关于指数函数的两个有用的性质是 $\exp(x_1+x_2)=\exp(x_1)\exp(x_2)$ 和 $\exp[c\cdot\log(x)]=x^c$。

A.5　微分学

在上节中，我们申述了几种基于微积分学的近似计算。令 $y=f(x)$，这里 f 表示某种函数。那么，对于 x 的微小变化，我们有

$$\Delta y\approx\frac{\mathrm{d}f}{\mathrm{d}x}\cdot\Delta x \tag{A.31}$$

式中，$\mathrm{d}f/\mathrm{d}x$ 为在起始点 x_0 计算的函数 f 的导数。我们还可把它写为 $\mathrm{d}y/\mathrm{d}x$。

例如，若 $y=\log(x)$，则 $\mathrm{d}y/\mathrm{d}x=1/x$。利用式(A.31)以及在 x_0 处计算的 $\mathrm{d}y/\mathrm{d}x$，我们就有 $\Delta y\approx(1/x_0)\Delta x$ 或 $\Delta\log(x)\approx\Delta x/x_0$，这就是式(A.22)所给的近似计算。

在应用计量经济学时，回顾一下为数不多的一些函数的导数是有所助益

的，因为我们要用导数来定义一个函数在给定点的斜率。这样我们就能利用式（A.31）来求当 x 有微小变化时 y 的近似变化。对于线性函数，导数正是我们所盼望的直线的斜率：若 $y=\beta_0+\beta_1 x$，则 $\mathrm{d}y/\mathrm{d}x=\beta_1$。

若 $y=x^c$，则 $\mathrm{d}y/\mathrm{d}x=cx^{c-1}$。两个函数之和的导数就是两个导函数之和 $\mathrm{d}[f(x)+g(x)]/\mathrm{d}x=\mathrm{d}f(x)/\mathrm{d}x+\mathrm{d}g(x)/\mathrm{d}x$。任何函数的常数倍（数）的导数等于该函数的导数的同样常数倍：$\mathrm{d}[cf(x)]/\mathrm{d}x=c[\mathrm{d}f(x)/\mathrm{d}x]$。这些简单的法则使我们能求出更复杂的函数的导数。其他法则诸如乘积、商以及链法则对于学过微积分的人来说应是他们所熟悉的，我们就不在这里复习了。

659 经济学中常用的一些函数及其导（函）数是：

$$y=\beta_0+\beta_1 x+\beta_2 x^2;\ \mathrm{d}y/\mathrm{d}x=\beta_1+2\beta_2 x$$

$$y=\beta_0+\beta_1/x;\mathrm{d}y/\mathrm{d}x=-\beta_1/(x^2)$$

$$y=\beta_0+\beta_1\sqrt{x};\ \mathrm{d}y/\mathrm{d}x=(1/2)x^{-1/2}$$

$$y=\beta_0+\beta_1\log(x);\ \mathrm{d}y/\mathrm{d}x=\beta_1/x$$

$$y=\exp(\beta_0+\beta_1 x);\ \mathrm{d}y/\mathrm{d}x=\beta_1\exp(\beta_0+\beta_1 x)$$

在最后的一个表达式中如果 $\beta_0=0$ 且 $\beta_1=1$，便得到 $y=\exp(x)$，同时 $\mathrm{d}y/\mathrm{d}x=\exp(x)$。

在 A.4 节中，我们看到在使用微积分时，方程（A.26）定义一个常弹性模型。弹性的微积分定义是 $\frac{\mathrm{d}y}{\mathrm{d}x}\cdot\frac{x}{y}$。可以证明，当方程（A.26）成立时，利用对数和指数函数的性质，$\frac{\mathrm{d}y}{\mathrm{d}x}\cdot\frac{x}{y}=\beta_1$。

当 y 为多变量的函数时，**偏导数**（partial derivative）的概念是重要的。假定

$$y=f(x_1,x_2) \tag{A.32}$$

就有两个偏导数，一个关于 x_1，另一个关于 x_2。y 对 x_1 的偏导数，这里记为 $\frac{\partial y}{\partial x_1}$，就是把 x_2 看做常数时式（A.32）对 x_1 的普通导数。类似地，$\frac{\partial y}{\partial x_2}$ 就是把 x_1 固定时式（A.32）对 x_2 的导数。

偏导数之所以有用就和普通导数有用的道理差不多。我们可以把 y 的变化近似计算为

$$\Delta y\approx\frac{\partial y}{\partial x_1}\cdot\Delta x_1,\text{固定 } x_2 \text{ 不变} \tag{A.33}$$

按照这个方法，和我们在线性模型中所做的那样，微分学能使我们在非线性模型中定义的**偏效应**（partial effect）。事实上，如果

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2$$

则

$$\frac{\partial y}{\partial x_1}=\beta_1,\quad \frac{\partial y}{\partial x_2}=\beta_2$$

这些偏导数可被视为 A.2 节中所定义的偏效应。

一个更复杂一点的例子是

$$y = 5 + 4x_1 + x_1^2 - 3x_2 + 7x_1 \cdot x_2 \qquad (A.34)$$

现在，式（A.34）对 x_1 的（偏）导数（视 x_2 为常数）就是

$$\frac{\partial y}{\partial x_1} = 4 + 2x_1 + 7x_2$$

660 要注意该导数是怎样地依赖于 x_1 和 x_2 的，式（A.34）对 x_2 的（偏）导数则是 $\frac{\partial y}{\partial x_2} = -3 + 7x_1$ ，它仅依赖于 x_1。

例 A.8　有交互作用的工资方程

把工资同受教育年数和工作经验（以年计）联系起来的一个函数是

$$\begin{aligned} wage = {} & 3.10 + 0.41educ + 0.19exper - 0.004exper^2 \\ & + 0.007educ \cdot exper \end{aligned} \qquad (A.35)$$

经验对工资的偏效应就是式（A.35）对经验的偏导数：

$$\frac{\partial wage}{\partial exper} = 0.19 - 0.008exper + 0.007educ$$

这是增加一年工作经验所导致的工资近似变化。注意这个偏效应与 exper 和 educ 的初始水平有关。例如，一个从 $educ = 12$ 和 $exper = 5$ 开始的工人，再增加一年工作经验，将使每小时工资增加约 $0.19 - 0.008(5) + 0.007(12) \doteq 0.234$ 或 23.4 美分。准确的变化可以通过在点（$exper = 5$，$educ = 12$）处和在点（$exper = 6$，$educ = 12$）处计算式（A.35），然后取二者之差而得到。计算结果是 0.23，这和近似计算的结果非常相近。

在求单变量或多变量函数的最小化或最大化时，微分计算起着重要的作用。如果 $f(x_1, x_2, \cdots, x_k)$ 是 k 个变量的一个可微函数，则 x_1^*，x_2^*，…，x_k^* 在 x_j 的所有可能值域上最小化或最大化 f 的必要条件是

$$\frac{\partial f}{\partial x_1}(x_1^*, x_2^*, \cdots, x_k^*) = 0, \quad j = 1, 2, \cdots, k \qquad (A.36)$$

换言之，在 x_k^* 处计算的 f 的全部偏导数必须为零。这些条件被叫做函数最小化或最大化的一阶条件。实践中我们希望解出方程（A.36）的 x_k^*，然后利用别的准则以判断我们是否已得到函数的最小值或最大值。这里，我们不需要讨论这些准则。[关于多变量微积分及其在优化函数中的应用，可参考 Sydsaeter 和 Hammond（1995）。]

▶ 小 结

这里所检验的数学工具，对于理解回归分析是至关重要的。概率和统计将在附录B和C中讨论。非线性函数——特别是二次、对数和指数函数 661 ——的材料，对理解现代应用经济研究十分重要。虽然某些数学推导要用到微积分，但对这些函数所要求的掌握程度并不涉及深入的微积分知识。

关键术语

(平)均值
其他情况不变
(恒)常弹性模型
导(函)数
描述统计量
递减边际效应
弹性
指数函数
截距
线性函数
对数函数
边际效应
中位数
自然对数
非线性函数
偏导数
偏效应
百分数/比变化
百分点变化
比例变化
半弹性
斜率
总和运算子

习 题

A.1 下表含有10个家庭的每月住房开支。

家庭	每月住房开支 （美元）
1	300
2	400
3	350
4	1 100
5	640
6	480
7	450
8	700
9	670
10	530

662 (i) 求每月住房开支的均值。

(ii) 求每月住房开支的中位数。

(iii) 如果每月住房开支以百美元计而不是以美元计，平均开支和中位数开支会是什么？

(iv) 假使第 8 号家庭的每月住房开支增大到 900 美元，而其余家庭的开支不变，再计算住房开支的均值和中位数。

A.2 假令下列方程描述某学期的平均旷课次数（*missed*）和到学校的距离（*distance*，以公里计）之间的关系：

$$missed = 3 + 0.2\, distance$$

(i) 画出这条直线并标明坐标，你怎样解释这个方程的截距？

(ii) 某人住处离学校 5 公里，他的旷课平均次数是多少？

(iii) 住在离学校 10 公里的地方和住在离学校 20 公里的地方，其平均旷课次数相差多少？

A.3 在例 A.2 中，音碟的需求量同价格和收入的关系是 $quantity = 120 - 9.8price + 0.3income$。如果 $price = 15$，而 $income = 200$，问对音碟的需求是什么？这个答案对于用线性函数描述需求曲线会有什么提示？

A.4 假定美国的失业率从某年的 6.4%变到其下一年的 5.6%。

(i) 失业率的百分点下降了多少？

(ii) 失业率下降了多大的一个百分数？

A.5 假定持有某企业股票的回报从某年的 15%变到下年的 18%。大多数股民声称股票回报只增加了 3%，而首席执行官则声称该企业股票的回报已上升 20%。试解释其中的分歧。

A.6 假定 A 某年挣 35 000 美元，而 B 某年挣 42 000 美元。

(i) 求 B 某薪水超过 A 某薪水的准确百分数。

(ii) 利用自然对数之差求近似的百分数差异。

A.7 假定以下模型描述年薪（*salary*）与先前参加劳动市场的年数（即经验 *exper*）之间的关系：

$$\log(salary) = 10.6 + 0.027exper$$

(i) 当 $exper = 0$ 时 *salary* 如何？当 $exper = 5$ 时又如何？[提示：你需要作指数化运算。]

663 (ii)利用式（A.28）计算当 *exper* 增加 5 年时年薪的百分数增加的近似值。

(iii) 利用 (i) 的结果计算当 $exper = 5$ 和 $exper = 0$ 时的准确百分数差异。

A.8 令 *grthemp* 代表 1990—1995 年县一级的就业的比例增长，并令 *salestax* 代表以比例形式给出的县销售税率，试解释下列方程中的截距和斜率：

$$grthemp = 0.043 - 0.78salestax$$

A.9 假定某种农作物的产量（蒲式耳/每亩）与施肥量（磅/每亩）的关系为

$$yield = 120 + 0.19\sqrt{fertilizer}$$

(i) 用几个数值代入 *fertilizer*，勾画出这个关系式的图形。

(ii) 把这个图形同 *yield* 和 *fertilizer* 之间的线性函数相比较。

附录 B　概率论基本知识

664 本附录概括了基础概率的重点概念。附录 B 和 C 主要属于复习性质，并不想用来代替一门概率与统计课程。然而，凡是本书用到的概率和统计学概念，全都包括在这些附录中了。

概率本身就是工商企业、经济学以及其他社会科学学生所感兴趣的。例如，一个航空公司要考虑只有 100 个座位的航班应接受多少个预订座位的问题。如果少于 100 人需要预订座位，就应该全部接受。但若多于 100 人要求预订怎么办？一个稳妥的办法就是最多接受 100 个预订。然而，有些人预订了座位，到起航时却不见人来。所以，即使接受了 100 个预订，仍存在飞机未能满载的机会，结果是造成了公司收入的损失。另一个策略是接受多于 100 个预订，并希望到时有些人不来，使得最后的乘客数尽可能接近 100 人。但由于订座过多，这一政策将冒要对搭不上飞机的人进行赔偿的风险。

在这种情况中的一个自然的问题是：公司接受多少预订座位才是最优的？这并非一个无关紧要的问题。不管怎样，给出一定的（关于航空公司成本和人们不会错过乘坐其订座的频率）信息，我们能利用基础概率知识获得一个解答。

B.1 随机变量及其概率分布

假使我们掷一枚钱币10次，计算共出现正面的次数，这就是一个**实验**(experiment)的例子。一般地说，一个实验是指，至少在理论上能够无限地重复下去的任何一种程序，并且它有一个良好定义的结果集。原则上，我们可以一次又一次地完成掷钱币的程序。在掷钱币之前，我们知道出现正面的次数是一个0~10的整数，因此这个实验的结果是良好定义的。

一个**随机变量**(random variable)是指一个具有数值特征并由一个实验来决定其结果值的变量。在掷钱币的例子中，掷一枚钱币10次出现正面的次数就是随机变量的一个例子。在掷币10次之前，我们不知道会出现正面多少次。一旦掷完10次并统计正面次数时，我们就得到随机变量在这个实验的一次特定试验中的结果。再做一次试验可能会产生不同的结果。

665 在前面提到的飞机订座例子中，不错过预订航班的人数是一个随机变量：在特定航班起飞之前，我们不知道有多少人会出现。

在分析工商企业和社会科学中所收集的数据时，对随机变量及其性质有一个基本的了解是重要的。在整个附录B和C中，按照概率和统计学的惯例，我们一律用大写字母如常见的W，X，Y和Z表示随机变量，而用相应的小写字母w，x，y和z表示随机变量的特定结果。例如，在掷币实验中，令X为一枚钱币投掷10次出现正面的次数。所以，X并不是任何具体数值，但我们知道X将在集合$\{0,1,2,\cdots,10\}$中取一个值。比方说，一个特殊的结果$x=6$。

我们用下标表示一系列随机变量。例如，我们记录在美国随机选择的20个家庭的上年收入。可以用$X_1,X_2,\cdots,X_{20}$表示这些随机变量，并用$x_1,x_2,\cdots,x_{20}$表示其特殊结果。

如在定义中所表明的，随机变量总是定义为取有数值特征的变量，即使它们描述的是一些定性事件。例如，考虑只掷一枚钱币，其两个结果是正面和反面。我们可以定义一个随机变量如下：如果出现正面则$X=1$；如果出现反面则$X=0$。

一个只能取0和1两个值的随机变量叫做**贝努里**(Bernoulli)[**或二值**(binary)]**随机变量**。在基础概率中有一个传统，把事件$X=1$叫做“成功”，而事件$X=0$叫做“失败”。在具体应用中，“成功—失败”的称谓不一定符合通常的成功或失败概念，但它却是我们准备采用的一个有用术语。

离散随机变量

离散随机变量(discrete random variable)是指一个只取有限个或可数的无

限个值的随机变量。"可数的无限个"是这样一个概念：虽然随机变量可取无限个值，但这些值可以和正整数一一对应起来。由于"可数的无限个"和"不可数的无限个"之间的差别有些玄妙，我们将集中讨论只取有限个值的离散随机变量。拉森和马克斯（Larsen and Marx，1986，第 3 章）有详细的论述。

贝努里随机变量是离散随机变量的最简单例子。惟一需要用以完全描述贝努里随机变量的行为的，仅是它取值为 1 的概率。在掷币的例子中，如果这枚钱币是"公平"的，那么 $P(X=1)=1/2$（读作"X 等于 1 的概率是一半"）。由于概率的总和必须是 1，所以还有 $P(X=0)=1/2$。

社会科学家所感兴趣的事物不限于掷币，因此我们还须考虑更一般的情形。再次考虑航空公司必须决定让多少人预订只有 100 个座位的航班问题。666 这个问题可以用多个贝努里随机变量的方式分析如下：对一个随机挑选的乘客，定义这样一个贝努里随机变量 X，如果预订座位的人到时出现，则 $X=1$；否则 $X=0$。

没有理由认为某一顾客出现的概率是 1/2；原理上，这个概率可以是 0 与 1 之间的任何一个数。且把它叫做 θ，于是

$$P(X=1)=\theta \tag{B.1}$$

$$P(X=0)=1-\theta \tag{B.2}$$

例如，若 $\theta=0.75$，则有 75% 的机会乘客在订座之后到时出现，并有 25% 的机会到时不出现。从直观上看便知，θ 值的大小在决定航空公司的订座策略中起着关键作用。给定飞机订座的历史数据，怎样估计 θ 是数理统计学的一个课题，在附录 C 中我们要转到这类问题上。

更一般的情形是，一个离散随机变量要由它的全部可能值和取每一值的相应概率来完全描述。如果 X 取 k 个可能值 $\{x_1, x_2, \cdots, x_k\}$，则其概率 $p_1, p_2, \cdots, p_k$ 被定义为

$$p_j=P(X=x_j), \quad j=1,2,\cdots,k \tag{B.3}$$

其中每个 p_j 都在 0～1 之间，并且

$$p_1+p_2+\cdots+p_k=1 \tag{B.4}$$

方程（B.3）读作"X 取值 x_j 的概率等于 p_j"。

方程（B.1）和（B.2）表明一个贝努里随机变量的成功和失败概率完全由 θ 值决定。由于贝努里随机变量如此流行，我们就为它设计了一个特殊符号：$X \sim \text{Bernoulli}(\theta)$，读作"$X$ 有一个成功概率等于 θ 的贝努里分布"。

X 的**概率密度函数**（probability density function，pdf）概括了 X 的可能结果以及相应概率的信息：

$$f(x_j)=p_j, \quad j=1,2,\cdots,k \tag{B.5}$$

而且对某个 j 凡是不等于 x_j 的 x 都有 $f(x)=0$。换言之，对任何实数 x，$f(x)$ 都是随机变量 X 取该特定值 x 的概率。当我们涉及多于一个随机变量时，有时用得着对所考虑的 pdf 加一个下标：例如 f_x 是 X 的 pdf，f_y 是 Y 的

pdf，等等。

给定任一离散随机变量的 pdf，就不难计算关于该随机变量的任何事件的概率。例如，设 X 为一名篮球运动员在两次罚球中命中的次数。因此 X 的三个可能值是 $\{0, 1, 2\}$。假定 X 的 pdf 是

$$f(0)=0.20、f(1)=0.44，\quad f(2)=0.36$$

662 这三个概率之和必定是 1。利用这个 pdf，我们能算出该运动员至少投入一球的概率：$\mathrm{P}(X\geqslant1)=\mathrm{P}(X=1)+\mathrm{P}(X=2)=0.44+0.36=0.80$。$X$ 的 pdf 如图 B.1 所示。

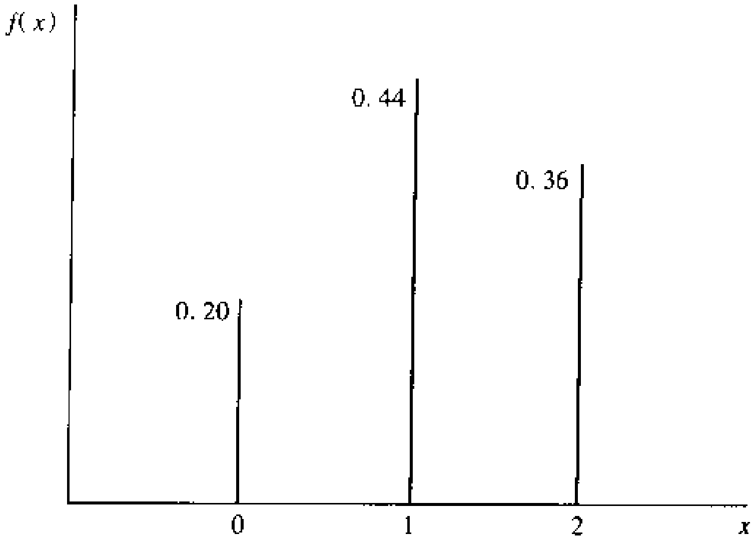


图 B.1　两次罚球命中次数的 pdf

连续随机变量

一个连续随机变量是指取任何实数的概率都是零的一个变量。这个定义有点违反直觉，因为在任何应用中，我们最终都会观测到一个随机变量取得的某种结果。这里的思想是，一个连续随机变量 X 的可能取值之多，我们无法用正整数去一一为它们计数，因而，为了逻辑上的无矛盾性，X 必须能以零的概率取每一个值。由于实际测量总是离散的，所以最好是把取众多值的随机变量看做是连续的。例如，一件商品的价格的最精细度量就是用美分来表示。我们可以想像按大小把价格的所有可能值都一一列出（即使这一排列可以无限地继续下去），从而技术上把价格变成了一个离散随机变量。然而，价格的可能值毕竟太多了，把它套到离散随机变量的操作方法中是不现实的。

我们可以对连续随机变量（简记随机变量为 rv）定义一个概率密度函数，并且和定义离散 rv 那样，pdf 提供了关于随机变量的可能结果的信息。但是，要讨论一个连续 rv 取某一特定值的概率是没有意义的，我们用一个

连续 rv 的 pdf 仅限于计算取值在一定范围内的事件的概率。例如，设 a 和 b 为常数，且 $a<b$。X 落在 a 与 b 之间的概率 $\mathrm{P}(a\le X\le b)$ 就是 pdf 曲线之
668 下 a 与 b 之间的面积，如图 B.2 所示。如果你熟悉微积分，你会认出它就是函数 f 在点 a 和 b 之间的积分。pdf 之下的整个面积必然等于 1

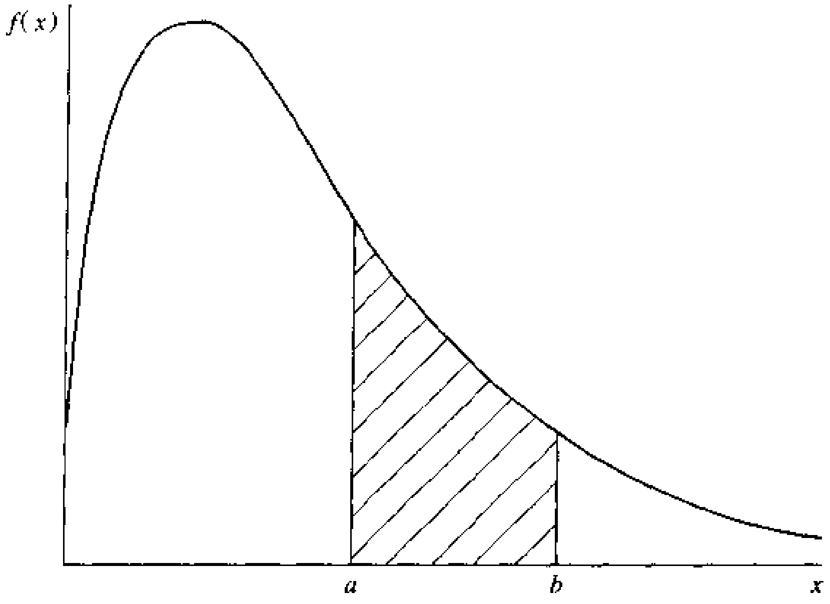


图 B.2　X 落在点 a 和 b 之间的概率

在计算连续随机变量的概率时，最方便使用**累积分布函数**（cumulative distribution function，cdf）。设 X 为任何随机变量，它的对任何实数 x 的 cdf 被定义为

$$F(x)=\mathrm{P}(X\le x) \tag{B.6}$$

对于离散随机变量，将 pdf 对所有 $x_j\le x$ 的 x_j 值求和就得到式（B.6）。而对于一个连续随机变量，$F(x)$ 就是概率密度函数 f 之下点 x 以左的面积。因为 $F(x)$ 就是一个概率，它总是在 0 与 1 之间。此外，如果 $x_1<x_2$，则 $\mathrm{P}(X\le x_1)\le \mathrm{P}(X\le x_2)$，即 $F(x_1)\le F(x_2)$。这意味着 cdf 是 x 的一个增（至少非减）函数。

cdf 对于计算概率有用的两个重要性质如下：

$$\mathrm{P}(X>c)=1-F(c)\text{，}c\text{ 为任意数} \tag{B.7}$$

$$\mathrm{P}(a<X\le b)=F(b)-F(a)\text{，}a<b \tag{B.8}$$

669 在计量经济学的学习中，我们仅用 cdf 来计算连续随机变量的概率，所以在表述概率的不等式中就不用区分是否严格不等。也就是说，对于一个连续随机变量 X，有

$$\mathrm{P}(X\ge c)=\mathrm{P}(X>c) \tag{B.9}$$

和

$$\mathrm{P}(a<X<b)=\mathrm{P}(a\le X\le b)=\mathrm{P}(a\le X<b)=\mathrm{P}(a<X\le b) \tag{B.10}$$

方程（B.9）和（B.10），连同方程（B.7）和（B.8），大大增加了利用连续cdf计算概率的机会。

人们已将概率和统计学中的所有重要连续分布的累积分布函数制成表格，其中最为人们熟知的是正态分布，我们将在B.5节中连同其他有关的一些分布加以介绍。

B.2 联合分布、条件分布与独立性

在经济学中，我们常对涉及多于一个随机变量的事件的发生感兴趣。例如，在先前的飞机订座例子中，航空公司也许对订座者到时是否出现同时是否公务旅客感兴趣；这就是联合分布的一个例子。或者，公司也许对如下的条件分布感兴趣：在某人是一位公务旅客的条件下，他或她到时出现的概率是什么？在下两节里，我们将把联合分布和条件分布的概念以及随机变量的独立性这一重要概念加以规范化。

联合分布与独立性

令X和Y为离散随机变量，那么(X,Y)的**联合分布**（joint distribution）由它们的联合概率密度函数完全描述：

$$f_{X,Y}(x,y)=\mathrm{P}(X=x,Y=y) \tag{B.11}$$

上式右端是$X=x$和$Y=y$的概率。当X和Y为连续时，我们又可以定义一个联合pdf，但由于本书并不明显使用连续随机变量的联合pdf，所以我们将不提供其中细节。

有一种情形，如果我们知道X的pdf和Y的pdf，就容易得到它们的联
670 合pdf。特别是，我们说X和Y是独立的，当且仅当，对一切x和y，有

$$f_{X,Y}(x,y)=f_X(x)f_Y(y) \tag{B.12}$$

式中，f_X为X的pdf；f_Y为Y的pdf。在多个随机变量的论述中，f_X和f_Y这两个pdf常称之为**边际概率密度函数**（marginal probability density function），以区别于联合pdff_{XY}。这里的独立性定义兼适用于离散和连续随机变量。

最容易弄明白式（B.12）的含义的方法是考虑离散的情形。如果X和Y都是离散的，那么式（B.12）就等同于

$$\mathrm{P}(X=x,Y=y)=\mathrm{P}(X=x)\mathrm{P}(Y=y) \tag{B.13}$$

换句话说，$X=x$且$Y=y$的概率就是两个概率$\mathrm{P}(X=x)$和$\mathrm{P}(Y=y)$之积。式（B.13）的一个含义是，联合概率是相当容易计算的，因为所需要知道的仅仅是$\mathrm{P}(X=x)$和$\mathrm{P}(Y=y)$。

如果两个随机变量不是独立的，就说它们是相依的。

例 B.1　罚球命中率

考虑篮球运动员的两次罚球。令 X 为贝努里随机变量：如果第一次命中等于 1，否则等于 0。再令 Y 为贝努里随机变量：如果第二次命中等于 1，否则等于 0。假使该运动员每次罚球的命中率都是 80%，即 $\mathrm{P}(X=1)=\mathrm{P}(Y=1)=0.8$，问两罚两中的概率是多少？

如果 X 和 Y 独立，就很容易回答这个问题：$\mathrm{P}(X=1, Y=1)=\mathrm{P}(X=1)\mathrm{P}(Y=1)=(0.8)(0.8)=0.64$。因此，有 64% 的机会两罚两中。如果第二次命中的机会依赖于第一次是否命中，也就是 X 和 Y 不独立，那么这种简单的计算是不正确的。

随机变量的独立性是一个异常重要的概念。在下一小节里，我们将阐明，如果 X 和 Y 独立，那么知道 X 的结果并不改变 Y 的各种可能结果的概率，反之亦然。关于独立性的一个有用的事实是，如果 X 和 Y 独立，而我们对任意函数 g 和 h 定义两个新的随机变量 $g(X)$ 和 $h(Y)$，则这些新随机变量也是独立的。

这里没有必要限于两个随机变量。如果 X_1，X_2，…，X_n 是 n 个随机变量，则它们的联合 pdf 是 $f(x_1, x_2, \cdots, x_n)=P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)$。随机变量 X_1，X_2，…，X_n 是独立的，当且仅当它们的联合 pdf 对任何（x_1，x_2，…，x_n）都是诸个别 pdf 的乘积。这一定义对连续随机变量来说也是对的。

独立性概念在推求概率和统计学中的一些经典分布时，扮演着重要角色。前面我们曾定义贝努里随机变量为表示某事件是否发生的 0—1 变量。我们常常对一系列贝努里试验中的成功次数感兴趣。独立贝努里试验的一个 671 标准例子是反复抛掷一枚钱币。由于任何特定的一次抛掷结果都和其他各次抛掷的结果不相干，所以作独立性假设是适宜的。

在更复杂的情况中，独立性也常常是一种合理的近似。在飞机订座的例子中，假使航空公司对某特定航班接受了 n 个预订。对于每一个 $i=1, 2, \cdots, n$，令 Y_i 为乘客到时是否出现的贝努里随机变量：如果乘客 i 到时出现，$Y_i=1$；如不然，$Y_i=0$。再次令 θ 表示成功（利用了订座）的概率，每一 Y_i 都有一个贝努里（θ）分布。作为一种近似，我们不妨假定这些 Y_i 是互相独立的，尽管实际上并非确切如此：一些人们会结队而行，这就意味着其中一人是否出现和其他所有人是否出现并非无关。然而，建立这类相关性的模型是复杂的，所以我们也许还愿意使用独立性作为一种近似方法。

我们主要关注的变量是在 n 个订座中有乘客出现的总数。把这个变量叫做 X。令每个 Y_i 取值 1，如果第 i 个订座人出现，就可以写 $X=Y_1+Y_2+\cdots+Y_n$。现假定每个 Y_i 有成功的概率 θ，并且这些 Y_i 是独立的，就可以证明 X 有一个**二项分布**（binomial distribution）。也就是，X 的概率密度函数是

$$f(x)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\quad x=0,1,2,\cdots,n \qquad \text{(B.14)}$$

式中，$\binom{n}{x}=\frac{n!}{x!\,(n-x)!}$，并且对任意整数 n，$n!$（读作"n 阶乘"）定义为 $n!=n\cdot(n-1)\cdot(n-2)\cdots1$。由惯例，$0!=1$。当一个随机变量 X 的 pdf 由式（B.14）给出时，我们写 $X\sim\text{Binomial}(n,\theta)$。方程（B.14）可用来对任何从 0 到 n 的 x 值计算 $P(X=x)$。

如果航班有 100 个座位，则航空公司关注 P（$X>100$）。假使在开始时 $n=120$，即公司接受了 120 个预订座位，而每个人出现的概率是 $\theta=0.80$，那么，$P(X>100)=P(X=101)+P(X=102)+\cdots+P(X=120)$。总和中的每一项概率都可从 $n=120$ 和 $\theta=0.80$ 的方程（B.14）求出，只须用适当的 x（从 101 到 120）代入便可。这是一项繁重的手算过程，幸而有许多统计软件包含有计算这类概率的命令。在本例中，有多于 100 个人出现的概率约为 0.659，这也许是航空公司不愿意承受的过多订座的风险。如果把预订数改为 110，那么多于 100 位乘客出现的概率则仅约为 0.024。

条件分布

在计量经济学中，我们感兴趣的常常是一个随机变量，且把它叫做 Y，是怎样和另外一个或多个随机变量联系起来的。暂时假定我们只对一个叫做 X 的变量的效应感兴趣。我们所能知道的关于 X 是怎样影响 Y 的，都包含在给定 X 时 Y 的**条件分布**（condition distribution）中，这一信息又总结为条件概率密度函数，其定义为：对所有 $f_X(x)>0$ 的 x 值，有

$$f_{Y|X}(y|x)=f_{X,Y}(x,y)/f_X(x) \qquad \text{(B.15)}$$

672 当 X 和 Y 都是离散变量时，式（B.15）最容易解释。这时

$$f_{Y|X}(y|x)=P(Y=y|X=x) \qquad \text{(B.16)}$$

上式右端读作"给定 $X=x$ 时 $Y=y$ 的概率"。当 Y 是连续变量时，由于前述理由，$f_{Y|X}(y|x)$不能直接解释为概率，但可通过计算条件 pdf 之下的面积来求出条件概率。

条件分布的一个重要特点是，如果 X 和 Y 是独立随机变量，知道 X 取什么值无助于确定 Y 取各值的概率（反之亦然）。这就是说，$f_{Y|X}(y|x)=F_Y(y)$，而且 $f_{X|Y}(x|y)=f_X(x)$。

例 B.2 罚球命中率

再次考虑篮球运动员两次罚球投篮的例子。假定条件密度是

$$f_{Y|X}(1|1)=0.85,\quad f_{Y|X}(0|1)=0.15$$
$$f_{Y|X}(1|0)=0.70,\quad f_{Y|X}(0|0)=0.30$$

这意味着球员第二次罚球命中的概率依赖于第一次罚球是否命中：如果第一次命中，则第二次命中的概率是0.85；如果第一次失误，则第二次命中的概率是0.70。这就是说，X 和 Y 不是独立的，而是相依的。

如果我们知道 $\mathrm{P}(X=1)$ 的话，则还可以计算 $\mathrm{P}(X=1,Y=1)$。且假定第一次命中的概率是0.8，即 $\mathrm{P}(X=1)=0.8$，那么，由式(B.15)我们得到两罚两中的概率为

$$\mathrm{P}(X=1,Y=1)=\mathrm{P}(Y=1|X=1)\cdot\mathrm{P}(X=1)=(0.85)(0.8)=0.68$$

B.3 概率分布的特征

为了许多目的，我们也只在少数几个方面对随机变量的分布感兴趣。所感兴趣的一些特征可分成三类：集中趋势的度量、变异或分散的度量以及两随机变量之间的关联性度量。我们将在B.4节讨论最后一类特征。

集中趋势的一种度量：期望值

期望值是在我们计量经济学学习中遇到的最重要概率性概念之一。设 X 为一随机变量。它的**期望值**（expected value 或 expectation），记作 $\mathrm{E}(X)$，并且有时记作 μ_X，甚至简单地记为 μ，就是对 X 的所有可能值的一个加权平均。673 权数由概率密度函数决定。有时期望值又被称为总体均值，特别是我们要强调 X 代表了总体中的某个变量之时。

当 X 是取有限个值，比方说 $\{x_1, x_2, \cdots, x_k\}$ 的离散随机变量时，期望值的准确定义最为简单。令 $f(x)$ 表示 X 的概率密度函数，则 X 的期望值为加权平均：

$$\mathrm{E}(X)=x_1f(x_1)+x_2f(x_2)+\cdots+x_kf(x_k)=\sum_{j=1}^{k}x_jf(x_j) \qquad (\mathrm{B}.17)$$

给定pdf在 X 的每一可能结果处的值，这是容易计算的。

例B.3 计算一个期望值

假定 X 分别以概率1/8，1/2和3/8取值−1，0和2，则

$$\mathrm{E}(X)=(-1)\cdot(1/8)+0\cdot(1/2)+2\cdot(3/8)=5/8$$

此例说明期望值的一个多少有点奇怪的现象：X 的期望值可以是一个甚至不是 X 的可能结果的数。我们知道 X 取的值是−1，0和2，但它的期

望值却是5/8。这使得期望值用来表征某些离散随机变量的集中趋势是有缺陷的，不过稍后我们将看到，期望值一类的计算可以是很有用的。

如果X是一个连续随机变量，则E(X)被定义为一个积分：

$$\mathrm{E}(X)=\int_{-\infty}^{\infty} x f(x)\mathrm{d}x \tag{B.18}$$

假定这个积分是良好定义的。这仍然可解释为一个加权平均。和离散情形不一样，E(X)总是X的可能结果的一个数。本书中，虽然我们要利用一些与特殊随机变量的期望值的概率有关的熟知成果，但我们并不需要利用积分去计算期望值。

给定随机变量X和函数$g(\cdot)$，我们可以制造一个新的随机变量$g(X)$。例如，若X是一随机变量，则X^2和$\log(X)(X>0)$也是随机变量。$g(X)$的期望值仍然是一个加权平均：

$$\mathrm{E}[g(X)]=\sum_{j=1}^{k} g(x_j)f_X(x_j) \tag{B.19}$$

或者对一个连续随机变量来说，有

$$\mathrm{E}[g(X)]=\int_{-\infty}^{\infty} g(x)f_X(x)\mathrm{d}x \tag{B.20}$$

例 B.4　X^2 的期望值

对于例B.3中的随机变量，令$g(X)=X^2$，便有

$$\mathrm{E}(X^2)=(-1)^2(1/8)+(0)^2(1/2)+(2)^2(3/8)=13/8$$

在例B.3中，我们算得$\mathrm{E}(X)=5/8$，因此$[\mathrm{E}(X)]^2=25/64$。这表明$\mathrm{E}(X^2)$和$[\mathrm{E}(X)]^2$不一样。事实上，对于一个非线性函数$g(X)$，$\mathrm{E}[g(X)]\neq g[\mathrm{E}(X)]$（除非是非常特殊的情形）。

如果X和Y是随机变量，则对任何函数g，$g(X,Y)$也是一随机变量，因而可以定义它的期望值。当X和Y都是离散的，分别取值$\{x_1, x_2, \cdots, x_k\}$和$\{y_1, y_2, \cdots, y_m\}$时，其期望值就是

$$\mathrm{E}[g(X,Y)]=\sum_{h=1}^{k}\sum_{j=1}^{m} g(x_h,y_j)f_{X,Y}(x_h,y_j)$$

式中，$f_{X,Y}$为(X, Y)的联合pdf。对于连续随机变量，这个定义因涉及积分要变得更复杂些，但这里我们不需要它。推广到多于两个随机变量是容易的。

期望值的性质

在计量经济学中我们并不那么关心计算各种分布的期望值；重要的计算

别人已做了多遍，我们尽可承认并加以利用。但我们需要用少数简单的规则去运算一些期望值。这些规则是如此重要，故予标识如下：

性质 E.1

$E(c)=c$，c 为任意常数

性质 E.2

$E(aX+b)=aE(X)+b$，a，b 为任意常数

性质 E.2 的一个有用的含义是，如果 $\mu=E(X)$，并且定义一个新随机变量 $Y=X-\mu$，那么，这相当于在 E.2 中取 $a=1$ 和 $b=-\mu$。

作为性质 E.2 的一个例子，令 X 为某地某日中午摄氏温度；假定预期温度为 $E(X)=25$。现令 Y 为华氏温度，则 $Y=32+(9/5)X$。由性质 E.2，华氏期望温度为 $E(Y)=32+(9/5)\cdot E(X)=32+(9/5)\cdot 25=77$。

广而言之，计算多个随机变量的线性函数的期望值是容易的。

675 **性质 E.3**

如果 $\{a_1, a_2, \cdots, a_n\}$ 是常数，而 $\{X_1, X_2, \cdots, X_n\}$ 是随机变量，则

$$E(a_1X_1+a_2X_2+\cdots+a_nX_n)=a_1E(X_1)+a_2E(X_2)+\cdots+a_nE(X_n)$$

或者利用总和符号，有

$$E\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_iE(X_i) \tag{B.21}$$

作为一个特例，取每个 $a_i=1$，我们有

$$E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i) \tag{B.22}$$

因此，总和的期望值就是期望值的总和。在数理统计的演算中常常用到这个性质。

例 B.5　求期望收入

令 X_1，X_2 和 X_3 分别为一馅饼店在某日出售的小、中、大馅饼的个数。这些随机变量的期望值是 $E(X_1)=25$、$E(X_2)=57$ 和 $E(X_3)=40$。小、中、大馅饼的价格是 5.50 美元、7.60 美元和 9.15 美元。因此，该日出售馅饼的期望收入为

$$\begin{aligned}&E(5.50X_1+7.60X_2+9.15X_3)\\&=5.50E(X_1)+7.60E(X_2)+9.15E(X_3)\\&=5.50(25)+7.60(57)+9.15(40)\\&=936.70\end{aligned}$$

即 936.70 美元。这不过是期望的收入，任何特定日的实际收入都会有所差异。

我们还可以利用性质 E.3 来证明，若 $X\sim \text{Binomial}(n,\theta)$，则 $E(X)=n\theta$。

即 n 次贝努里试验的期望成功次数不外是试验次数乘以任一次试验成功的概率。把 X 写成 $X = Y_1 + Y_2 + \cdots + Y_n$，其中 $Y_i \sim \text{Bernoulli}(\theta)$，就容易看出这点。这样一来，有

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(Y_i) = \sum_{i=1}^{n} \theta = n\theta$$

可以把这个结果应用到飞机订座的例子中去。在这个例子中，公司接受了 $n = 120$ 个订座，并且乘客出现的概率是 $\theta = 0.85$。乘客出现的期望数是 $120(0.85) = 102$。因此，如果现有 100 个座位，预期出现的乘客数就已过大，这关系到航空公司接受 120 个预订是否明智的问题。

其实，航空公司应该定义一个利润函数，用以说明每售出一个座位的净收入以及每把一位乘客从航班中排挤出去的成本。因为出现的人数是一随机变量，所以利润函数也是随机的。令 r 为来自每位乘客的利润。（为简单起见，你可把它看做机票的价格。）令 c 为每位被排挤出航班的获赔金额。r 和 c 都不是随机的，并假定对航空公司来说是已知的。令 Y 表示航班的利润。那么，对 100 个座位来说，有 676

$$\begin{aligned} Y &= rX, X \leqslant 100 \\ &= 100r - c(X - 100), X > 100 \end{aligned}$$

第一个方程给出当起飞前出现的乘客数不多于 100 时的利润；第二个方程则给出当出现的人数大于 100 时的利润。[在后一情形中，来自售票的纯收入是 $100r$，因为全部 100 个座位的机票均已售出，而 $c(X - 100)$ 就是接受多于 100 个订座的成本。] 利用 $X \sim \text{Binomial}\ (n,\ 0.85)$ 分布，其中 n 代表所接受的预订数，这个事实就可以求出作为 n（以及 r 和 c）的函数的期望利润 $\mathrm{E}(Y)$。要直接计算 $\mathrm{E}(Y)$ 将是很困难的，但可以通过计算机很快地求出。一旦给定 r 和 c，就能从不同的 n 值中寻搜最大化期望利润的 n 值。

集中趋势的另一种度量：中位数

用期望值定义一个随机变量的集中趋势只不过是可能方法之一，另一种度量集中趋势的方法是用**中位数**（median）。中位数的一般定义过于复杂，对我们来说还用不着。如果 X 是连续的，则中位数 m（比方说）就是这样一个数，pdf 之下的一半面积在 m 之左，另一半面积在 m 之右。

当 X 是离散的且取有限的奇数个值时，中位数就是按大小排序后居中的一个数。例如，若 X 可能取的值是 $\{-4, 0, 2, 8, 10, 13, 17\}$，则 X 的中位数是 8。若 X 可能取的是偶数个值，则实际上有两个中位数：有时取这两个数的平均，以得到惟一的一个中位数。例如，X 取的值是 $\{-5, 3, 9, 17\}$，则中位数是 3 和 9；如果取其平均，我们就得到中位数等于 6。

一般而言，中位数有时记为 $\mathrm{Med}(X)$，和期望值 $\mathrm{E}(X)$ 是不相同的。作为集中趋势的度量，不能说哪一个比另一个要好。

为了度量 X 的分布的中心，两者都是有效的方法。在一种特殊情形中，中位数和期望值（均值）是相同的。如果 X 的概率分布是对称于 μ 值而分布的，则 μ 既是期望值又是中位数。从数学上看，条件是对一切 x 都要有 $f(\mu+x)=f(\mu-x)$。图 B.3 说明了这种情形。

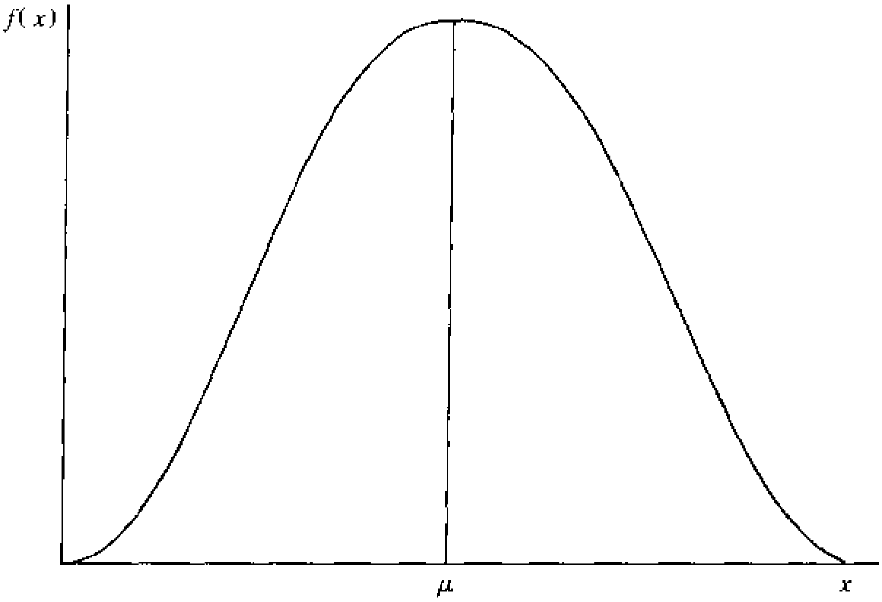


图 B.3　一个对称的概率分布

变异性的度量：方差与标准差

尽管集中趋势对一个随机变量提供了有价值的信息，它还不能告知我们关于这个随机变量的一切。图 B.4 展示了有相同均值的两个随机变量的 pdf。显然 X 的分布比 Y 的分布更紧密地集中在其中心的周围。我们自然希望有一种描述这种情形的简单方法。677

方差

对一个随机变量 X，令 $\mu=\mathrm{E}(X)$。为了度量 X 离开它的期望值多远，有多种方法，而最简单的一种代数方法是用差异的平方 $(X-\mu)^2$。（平方是为了消除距离度量的符号，以使最终得到的正值符合我们对距离的直观认识。）因这一距离随 X 的每一结果而变，故本身就是一个随机变量。正如我们需要用一个数来表征 X 的集中趋势那样，我们也需要用一个数来告知我们 X 平均而言离开 μ 多远。这样的一个数就是**方差**（variance），它告诉我们 X 对其均值的期望距离：

$$\mathrm{Var}(X)\equiv\mathrm{E}[(X-\mu)^2] \tag{B.23}$$

方差有时记为 σ_X^2，或在上下文中清楚时简记为 σ^2。由式（B.23），知方差必定是非负的。

作为一种计算方法，注意到下列关系是有用的：

$$\sigma^2 = \mathrm{E}(X^2 - 2X\mu + \mu^2) = \mathrm{E}(X^2) - 2\mu^2 + \mu^2 = \mathrm{E}(X^2) - \mu^2 \qquad (\mathrm{B}.24)$$

678

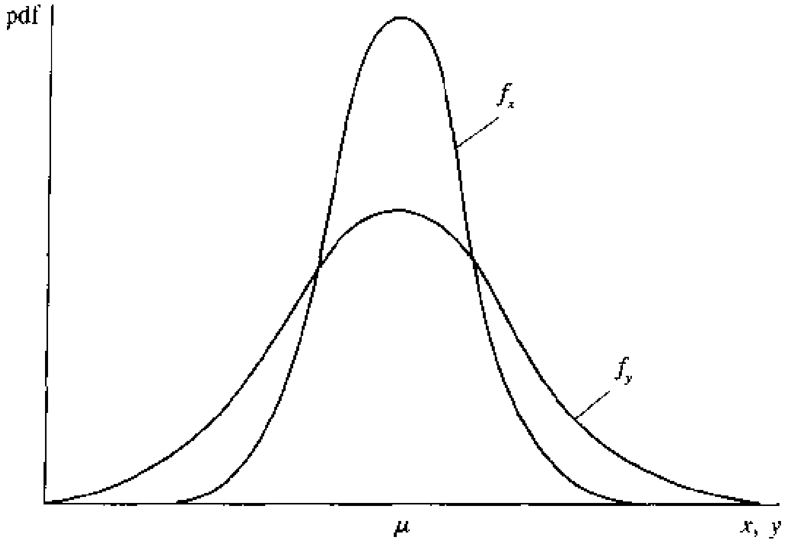


图 B.4 有相同均值但不相同分布的随机变量

利用式（B.23）或式（B.24）都没有区分连续或离散随机变量的必要：两种情形的方差定义都是一样的。最经常用的方法是，先计算E(X)，再计算 E(X^2)，然后利用式（B.24）。例如，若$X\sim$Bernoulli(θ)，则E(X) = θ，并且由于 $X^2 = X$，故 E(X^2) = θ。于是，由方程(B.24)知 Var(X) = E(X^2) − μ^2 = $\theta - \theta^2 = \theta(1-\theta)$。

方差的两个重要性质如下：

性质 VAR.1

当且仅当存在常数 c 使得P=($X = c$) = 1 时[这时 E(X) = c]，Var(X) = 0。

这个性质是说，任何常数的方差都是零，而且，如果一个随机变量有零方差，则它本质上就是常量。

性质 VAR.2

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X), \quad a, b \text{ 为任意常数}$$

679 这个性质的意思是，把一个常数加到一个随机变量上不会改变其方差，但用一个常数去乘一个随机变量将使其方差增大该常数的一个平方倍数。例如，若 X 指摄氏温度，而 $Y = 32 +$（9/5）X 为其华氏温度，则

$$\mathrm{Var}(Y) = (9/5)^2\mathrm{Var}(X) = (81/25)\mathrm{Var}(X)$$

标准差

一个随机变量的标准差，记为 sd（X），就是它的方差的正的平方根：

$$\mathrm{sd}(X) \equiv +\sqrt{\mathrm{Var}(X)}$$

标准差有时又记作 σ_X 或省去随机变量而记为 σ。标准差有两个性质可以直接从性质 VAR.1 和 VAR.2 推出。

性质 SD.1

$$\mathrm{sd}(c)=0,\quad c \text{ 为任意常数}$$

性质 SD.2

$$\mathrm{sd}(aX+b)=|a|\mathrm{sd}(X),\quad a,b \text{ 为任意常数}$$

特别是，如果 $a>0$，则

$$\mathrm{sd}(aX)=a\cdot\mathrm{sd}(X)$$

最后这个性质使标准差操作起来比方差更为自然。例如，假定 X 是以千美元度量的代表收入的一个随机变量。如果我们定义 $Y=1\,000X$，也就是说 Y 代表以美元度量的收入。假如 $\mathrm{E}(X)=20$，且 $\mathrm{sd}(X)=6$，那么 $\mathrm{E}(Y)=1\,000\mathrm{E}(X)=20\,000$，并且 $\mathrm{sd}(Y)=1\,000\mathrm{sd}(X)=6\,000$，可见期望值和标准差都增加相同的倍数 1 000。如果我们使用方差，将得到 $\mathrm{Var}(Y)=(1\,000)^2\mathrm{Var}(X)$，也就是 Y 的方差要比 X 的方差大 100 万倍。

标准化一个随机变量

作为方差和标准差的性质的一个应用——而且本身也是有实际意义的一个问题——假如给定随机变量 X，我们从它减去其均值 μ 并除以其标准差 σ，从而定义一个新的随机变量

$$Z \equiv \frac{X-\mu}{\sigma} \tag{B.25}$$

这又可写为 $Z=aX+b$，其中 $a\equiv(1/\sigma)$ 而 $b\equiv-(\mu/\sigma)$。于是，由性质 E.2，有

$$\mathrm{E}(Z)=a\mathrm{E}(X)+b=(\mu/\sigma)-(\mu/\sigma)=0$$

由性质 VAR.2，有

680

$$\mathrm{Var}(Z)=a^2\mathrm{Var}(X)=(\sigma^2/\sigma^2)=1$$

因此，随机变量 Z 有一个零均值和一个等于 1 的方差（或者标准差）。这一程序有时叫做将随机变量 X 标准化，而 Z 则叫做**标准化随机变量**（standardized random variable）。（在统计学入门教程中，有时把它叫做 X 的 z-变

换。）重要的是要记住在式（B.25）的分母中出现的是标准差而不是方差。我们即将看到，这一变换常用于统计推断。

作为特例，令$E(X)=2$和$\mathrm{Var}(X)=9$，则$Z=(X-2)/3$有零均值和单位方差。

B.4 联合与条件分布的特征

关联度：协方差与相关

虽然两个随机变量的联合 pdf 完整地描述了它们之间的关系，但仍用得着有描述它们平均看来是怎样互相变动的扼要度量手段。像期望值和方差那样，用一个数来概括整个分布的某个方面，现在要概括的是两个随机变量的联合 pdf。

协方差

令$\mu_X=E(X)$和$\mu_Y=E(Y)$，并考虑随机变量$(X-\mu_X)(Y-\mu_Y)$。如果X超过它的均值同时Y也超过它的均值，就有$(X-\mu_X)(Y-\mu_Y)>0$。如果$X<\mu_X$且$Y<\mu_Y$，那么仍然有这个结果。另一方面，如果$X>\mu_X$而$Y<\mu_Y$，或两个不等式都反过来，则$(X-\mu_X)(Y-\mu_Y)<0$。但这个乘积怎样能告诉我们X和Y之间有什么关系呢？

两个随机变量X和Y之间的**协方差**（covariance），有时叫做总体协方差，以强调它是关于描述一个总体的两个随机变量的，其定义为乘积$(X-\mu_X)(Y-\mu_Y)$的期望值：

$$\mathrm{Cov}(X,Y)\equiv E[(X-\mu_X)(Y-\mu_Y)] \qquad (B.26)$$

这有时又记为σ_{XY}。如果$\sigma_{XY}>0$，则平均而言，当X超过其均值时，Y也超过其均值。如果$\sigma_{XY}<0$，则平均而言，当X超过其均值时，Y低于其均值。

为了计算 Cov（X，Y），有几个有用的表达式如下：

$$\begin{aligned}\mathrm{Cov}(X,Y)&=E[(X-\mu_X)(Y-\mu_Y)]=E[(X-\mu_X)Y]\\&=E[X(Y-\mu_Y)]=E(XY)-\mu_X\mu_Y\end{aligned} \qquad (B.27)$$

由式（B.27）推知，如果$E(X)=0$或$E(Y)=0$，则$\mathrm{Cov}(X,Y)=E(XY)$。

681 协方差度量两个随机变量之间的**线性相依性**（linear dependence）。一个正的协方差表示两个随机变量同向移动，而一个负的协方差则表示两个随机变量反向移动。要解释一个协方差的大小，如我们即将看到的，可能有些

曲折。

由于协方差度量着两个随机变量是怎样一种关系，一个很自然的问题便是协方差和独立性概念是怎样联系起来的。这将由以下性质给出解答。

性质 COV.1

如果 X 和 Y 是独立的，则

$$\mathrm{Cov}(X,Y)=0$$

此性质来自方程（B.27）以及如果 X 和 Y 是独立的 $E(XY)=E(X)E(Y)$ 这一事实。重要的是要记住 Cov.1 反过来并不成立：X 和 Y 之间的零协方差并不意味着 X 和 Y 是独立的。事实上，存在有随机变量 X 和 $Y=X^2$ 使得 $\mathrm{Cov}(X,Y)=0$。［任何 $E(X)=0$ 且 $E(X^3)=0$ 的随机变量 X 都有此性质。］然而，如果 $Y=X^2$，则 X 和 Y 显然是不独立的：一旦知道了 X，必然知道 Y。所以，X 和 X^2 之间能够有零协方差看来颇为奇怪，而这正表明协方差作为随机变量之间的一般性的关联度量是有缺陷的。但当关系式至少近似于线性时，协方差是有用的。

协方差的第二个主要性质涉及线性函数之间的协方差。

性质 COV.2

对任意常数 a_1，b_1，a_2 和 b_2，有

$$\mathrm{Cov}(a_1X+b_1,a_2Y+b_2)=a_1a_2\mathrm{Cov}(X,Y) \qquad (B.28)$$

性质 COV.2 的一个重要含义在于，两个随机变量之间的协方差只须通过将两者或两者之一乘以一个常数倍，就能加以改变。这点在经济学中之所以重要是因为诸如货币变量、通货膨胀率均可通过不同的度量单位加以定义而不改变其实质。

最后，知道任何两个随机变量的协方差的绝对值必不超过它们的标准差的乘积是重要的；这一界限就是有名的柯西－施瓦兹不等式（Cauchy-Schwartz inequality）。

性质 COV.3

$$|\mathrm{Cov}(X,Y)|\leqslant \mathrm{sd}(X)\mathrm{sd}(Y)$$

相关系数

假定我们想知道劳动总体中受教育程度和年薪之间的关系，就可令 X 代表教育，Y 代表薪金，然后计算它们的协方差。然而我们得到的答案却依赖于我们对度量教育和薪金的选择。性质 COV.2 意味着教育和薪金之间的协方差视薪金以美元或以千美元度量而定，或者教育以月或以年计算而定。很明显，这些变量是怎样度量的对它们有多强的关系是没有影响的。但是，它们之间的协方差却与度量单位有关。 682

协方差与度量单位有关是它的一个缺陷，为克服这一缺陷，现引进 X

和 Y 的**相关系数**（correlation coefficient）：

$$\mathrm{Corr}(X,Y)\equiv\frac{\mathrm{Cov}(X,Y)}{\mathrm{sd}(X)\cdot\mathrm{sd}(Y)}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{B.29}$$

X 和 Y 的相关系数有时记作 ρ_{XY}（有时称为总体相关）。

因 σ_X 和 σ_Y 都是正数，$\mathrm{Cov}(X,Y)$ 和 $\mathrm{Corr}(X,Y)$ 必定有相同的符号，并且，当且仅当 $\mathrm{Cov}(X,Y)=0$ 时 $\mathrm{Corr}(X,Y)=0$。协方差的某些性质可移植到相关。如果 X 和 Y 是独立的，则 $\mathrm{Corr}(X,Y)=0$，但零相关不意味着是否独立。（相关系数也是线性相依的一个度量。）然而，由于下述性质，相关系数的大小比协方差的大小易于解释。

性质 CORR.1

$$-1\leqslant\mathrm{Corr}(X,Y)\leqslant 1$$

如果$\mathrm{Corr}(X,Y)=0$，或等价地 $\mathrm{Cov}(X,Y)=0$，则在 X 和 Y 之间无线性关系，并称 X 和 Y **不相关**（uncorrelated）；否则 X 和 Y 就是相关的。$\mathrm{Corr}(X,Y)=1$ 意味着一个完全的正线性关系，意思是说，我们对某常数 a 和某常数 $b>0$ 可以写 $Y=a+bY$。$\mathrm{Corr}(X,Y)=-1$ 则意味着一个完全的负的关系，使得对某个 $b<0$ 有 $Y=a+bX$。$+1$ 和 -1 两个极端情形很少出现。接近 1 或 -1 的 ρ_{XY} 值就表明了较强的线性关系。

如前面所得到的，X 和 Y 之间的相关不因 X 或 Y 的度量单位而变。下面说明了更一般的情形。

性质 CORR.2

对于常数 a_1，b_1，a_2 和 b_2，如果 $a_1a_2>0$，则

$$\mathrm{Corr}(a_1X+b_1,a_2Y+b_2)=\mathrm{Corr}(X,Y)$$

如果 $a_1a_2<0$，则

$$\mathrm{Corr}(a_1X+b_1,a_2Y+b_2)=-\mathrm{Corr}(X,Y)$$

作为一个例子，假定薪金和教育的总体相关是 0.15。这一度量将与用美元、千美元或任何其他单位计算薪金都无关；也不管用年、季、月或其他单位来衡量受教育的时间。

随机变量之和的方差

683 一旦定义了协方差和相关，就可以把方差的主要性质全部罗列出来。

性质 VAR.3

对于常数 a 和 b，有

$$\mathrm{Var}(aX+bY)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)+2ab\mathrm{Cov}(X,Y)$$

由此立即推知，如果 X 和 Y 是不相关的——从而 $\mathrm{Cov}(X,Y)=0$——则

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y) \tag{B.30}$$

和 $$\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y) \tag{B.31}$$

在后一情形中，要注意怎么会差的方差竟是（两）方差之和，而不是（两）方差之差。

作为式（B.30）的一个例子，令 X 为星期五夜晚某酒店赚到的利润，而 Y 为接着的星期六夜晚赚到的利润。因此 $Z=X+Y$，就是这两个夜晚赚的利润。假定 X 和 Y 都有一个 300 美元的期望值和一个 15 美元的标准差（因而方差为 225）。两夜晚的期望利润将是$\mathrm{E}(Z)=\mathrm{E}(X)+\mathrm{E}(Y)=2\cdot(300)=600$ 美元。如果 X 和 Y 是独立的，从而它们是不相关的，则总利润的方差是两方差之和：$\mathrm{Var}(Z)=\mathrm{Var}(X)+\mathrm{Var}(Y)=2\cdot(225)=450$。由此推知，总利润的标准差是$\sqrt{450}$或约为 21.21 美元。

表达式（B.30）和（B.31）可推广到多于两个变量的情形。为了阐明这一推广情形，我们需要一个定义。如果集合中的每一变量都与集合中每另一变量不相关，我们说随机变量是**两两不相关随机变量**（pairwise uncorrelated random variables）。就是说，对所有的$i\neq j$，有

$$\mathrm{Cov}(X_i,X_j)=0$$

性质 VAR.4

如果$\{X_1,\cdots,X_n\}$是两两不相关的随机变量且$\{a_i: i=1,\cdots,n\}$是常数，则

$$\mathrm{Var}(a_1X_1+\cdots+a_nX_n)=a_1^2\mathrm{Var}(X_1)+\cdots+a_n^2\mathrm{Var}(X_n)$$

用总和符号又可写为

$$\mathrm{Var}\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_i^2\mathrm{Var}(X_i) \tag{B.32}$$

性质 VAR.4 的一个特例是对所有的 i 取 $a_i=1$。这时，对两两不相关的随机变量来说，和的方差就是方差的和：

$$\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{Var}(X_i) \tag{B.33}$$

因为独立随机变量是不相关的（见性质 COV.1），故独立随机变量之和的方差是诸方差之和。

684 如果诸 X_i 不是两两不相关的，则$\mathrm{Var}\left(\sum_{i=1}^{n}a_iX_i\right)$的表达式要复杂得多；它不仅依赖于每一方差，还依赖于每一协方差。为了我们的目的，我们还不需要这个更一般的公式。

我们可以利用式（B.33）导出一个二项随机变量的方差。令 $X\sim\mathrm{Binomial}(n,\theta)$并写 $X=Y_1+\cdots+Y_n$，其中 Y_i 是独立 Bernoulli(θ)变量。于是，由式(B.33)，有

$$\mathrm{Var}(X)=\mathrm{Var}(Y_1)+\cdots+\mathrm{Var}(Y_n)=n\theta(1-\theta)$$

在航空订座的例子中 $n=120$ 和 $\theta=0.85$，那么到时出现的乘客人数的

方差是$120(0.85)(0.15) = 15.3$，因而它的标准差约为3.9。

条件期望

协方差和相关都是对两个随机变量之间的线性关系的度量，并且对称地处理两者。在社会科学中更多的情况是，我们想用另一个变量X去解释另一个变量Y。而且，如果Y和X有非线性形式的关系，则我们还希望知道这个形式。把Y叫做被解释（explained）变量，而X叫做解释（explanatory）变量。例如Y代表每小时工资，而X代表受过正式教育的年数。

我们曾介绍过给定X的Y的条件概率密度函数这个概念。这样，我们也许想知道工资的分布是怎样随教育水平而变化的。然而，我们常常需要用一个简单的方法来概括这个分布。由于在给定$X = x$下，Y的分布一般地说要依赖于这个x值，所以只用单一个数去概括是不够的。不过，我们可以通过给定X的Y的**条件期望**（conditional expectation），这有时又称条件均值，来概括Y和X的关系。意思是说，一旦我们知道X取了某个特定值x，我们就能根据X的这个结果算出Y的期望值。我们把这个期望值记作$\mathrm{E}(Y|X = x)$或简记$\mathrm{E}(Y|x)$。一般情形是，随着x的改变，$\mathrm{E}(Y|x)$也在改变。

当Y是取值$\{y_1, \cdots, y_m\}$的离散随机变量时，有

$$\mathrm{E}(Y|x) = \sum_{j=1}^{m} y_j f_{Y|X}(y_j|x)$$

当Y为连续时，$\mathrm{E}(Y|x)$由$y f_{Y|X}(y|x)$对y的所有可能值求积分来定义。好像无条件期望那样，条件期望也是一个对Y的全部可能值的加权平均，只不过这时的权数反映了X已取了某个特殊值的情形。因此，$\mathrm{E}(Y|x)$是x的某个函数，这个函数告诉我们Y的期望值是怎样地随x而变化的。

作为一个例子，令(X, Y)代表全体劳动者的一个总体，其中X为受教育年数，Y为每小时工资。那么，$\mathrm{E}(Y|X = 12)$就是总体中所有受了12年教育（相当于读完中学）的人的平均每小时工资。$\mathrm{E}(Y|X = 16)$则是所有受过16年教育的人的平均每小时工资。跟踪各种教育水平的期望值，对于工资和教育是怎样一种关系提供重要信息。参看图B.5中的图解。

685 原理上，可以在每一教育水平上求出每小时工资的期望值，然后将这些期望值列表。由于教育的变化范围很大——且可度量为一年的某个分数——用这种方法来表示平均工资和教育程度之间的关系是很繁琐的。计量经济学中的典型方法是设定一些足以刻画这种关系的简单函数。作为一个例子，假使给定$EDUC$的$WAGE$的期望值是如下的线性函数：

$$\mathrm{E}(WAGE|EDUC) = 1.05 + 0.45EDUC$$

假定这一关系对劳动者的总体成立，则有8年教育的人的平均工资是$1.05 + 0.45(8) = 4.65$或4.65美元。有16年教育的人的平均工资是8.25或8.25美

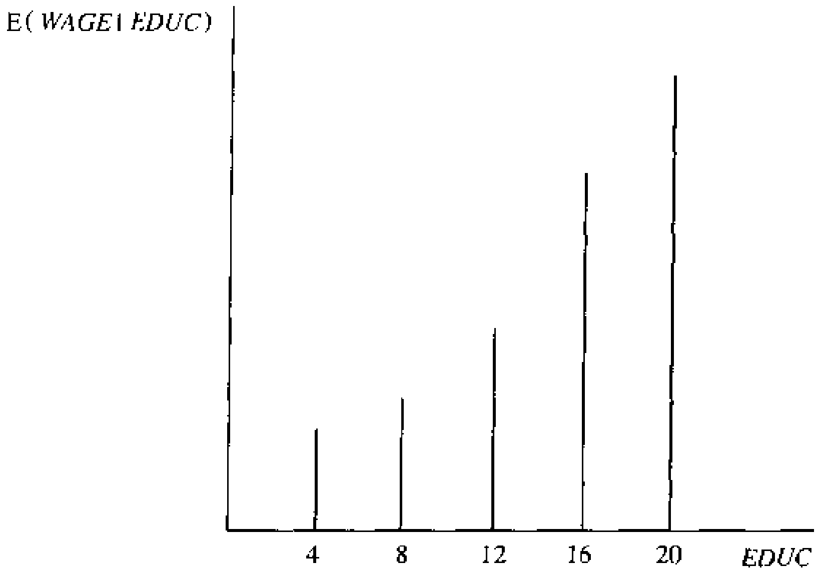


图 B.5　给定各种教育水平的每小时工资的期望值

元。$EDUC$ 的系数意味着每增加一年教育、平均每小时工资增加 0.45 美元或 45 美分。

条件期望也可能是个非线性函数。例如，假令 $\mathrm{E}(Y|x)=10/x$，其中 X 是一个恒大于零的随机变量。这个函数的图形见图 B.6。它可以代表一个需求函数，其中 Y 为需求量，而 X 为价格。如果 Y 和 X 的关系确实如此，则诸如相关分析一类的线性关联性分析是不适宜的。

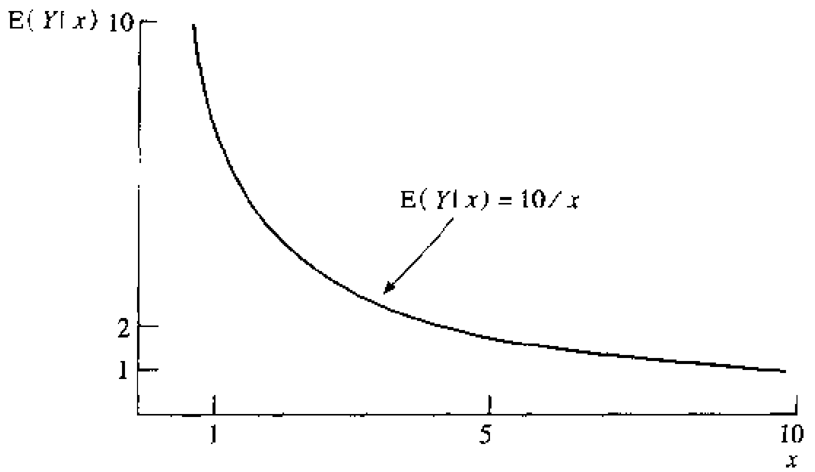


图 B.6　$\mathrm{E}(Y|x)=10/x$ 的图形

条件期望的性质

条件期望（CE）的一些基本性质对计量经济学分析中的演算、推导是有用的。

性质 CE.1

对任意函数 $c(X)$，有

$$E[c(X)|X]=c(X)$$

这个性质是说，当我们计算以 X 为条件的期望值时，X 的函数可视同常数。例如 $E(X^2|X)=X^2$。直观上，这无非是说，如果知道了 X，也就知道了 X^2。

性质 CE.2

对函数 $a(X)$和 $b(X)$，有

$$E[a(X)Y+b(X)|X]=a(X)E(Y|X)+b(X)$$

例如，我们能容易地计算像 $XY+2X^2$ 这样的一个函数的条件期望：

$$E(XY+2X^2|X)=XE(Y|X)+2X^2$$

下一个性质把独立性和条件期望的概念紧密联系起来。

性质 CE.3

如果 X 和 Y 是独立的，则

$$E(Y|X)=E(Y)$$

687 这个性质是说，如果 X 和 Y 是独立的，则在给定 X 时 Y 的期望值与 X 无关，这时$E(Y|X)$必定等于 Y 的（无条件）期望值。在工资与教育一例中，假使工资独立于教育，那么高中毕业生和大学毕业生的平均工资就会是一样的。因为这几乎无疑是错误的，所以我们不能假定工资与教育是独立的。

下面是性质 CE.3 的一个特例：如果 U 和 X 独立，并且$E(U)=0$，则

$$E(U|X)=0$$

条件期望还有一些性质，不免和$E(Y|X)$事实上是 X 的一个函数，比方说 $E(Y|X)$和 $\mu(X)$有关。由于 X 是一个随机变量，$\mu(X)$也就是一个随机变量。而且，$\mu(X)$有一个概率分布，因而有一个期望值。一般地说，要直接计算 $\mu(X)$的期望值，是十分困难的。**重期望律**(law of iterated expectation)告诉我们 $\mu(X)$的期望值就等于 Y 的期望值。我们把它写作：

性质 CE.4

$$E[E(Y|X)]=E(Y)$$

这个性质乍看起来有点难以掌握。它是说，如果我们先把$E(Y|X)$看做 X 的函数，再求这个函数的期望值，那么结果就是。这点并不显而易见，但可借助期望值的定义把它推导出来。

令 $Y=WAGE$ 和$X=EDUE$，其中 $WAGE$ 为每小时工资，而 $EDUC$ 为受教育年数。假定给定 $EDUC$ 的 $WAGE$ 期望值是$E(WAGE/EDUC)=4+0.60EDUC$，而且 $E(EDUE)=11.5$。那么，重期望律意味着$E(WAGE)=E(4+0.60EDUC)=4+0.60E(EDUE)=4+0.60(11.5)=10.90$ 或 10.90 美元/小时。

下一个性质是重期望律的一个更一般的表述。

性质 CE.4′

$$E(Y|X)=E[E(Y|X,Z)|X]$$

换句话说，求 $E(Y|X)$ 可分为两步。首先对任一其他的随机变量 Z 求 $E(Y|X,Z)$。然后求以 X 为条件的 $E(Y|X,Z)$ 的期望值。

性质 CE.5

如果 $E(Y|X)=E(Y)$，则 $\mathrm{Cov}(X,Y)=0$（因而 $\mathrm{Corr}(X,Y)=0$）。事实上，X 的每一函数都与 Y 不相关。

该性质的含义是，如果对 X 的知识不改变 Y 的期望值，则 X 和 Y 必须是不相关的。性质 CE.5 的逆并不成立：如果 X 和 Y 不相关，$E(Y|X)$ 仍可能依赖于 X。例如，取 $Y=X^2$，于是 $E(Y|X)=X^2$，这明明是 X 的一个函数。然而，正如我们在讨论协方差和相关时提到过的，X 和 X^2 可能不相关。688 条件期望抓住了相关分析所完全忽略掉的 X 和 Y 之间的非线性关系。

性质 CE.4 和 CE.5 有两个主要含义：如果 U 和 X 是这样的随机变量：$E(U|X)=0$，那么 $E(U)=0$ 而且 U 和 X 不相关。

性质 CE.6

如果 $E(Y^2)<\infty$ 并且对某个函数 g 有 $E[g(X)]^2<\infty$，则

$$E\{[Y-\mu(X)]^2|X\}\leqslant E\{[Y-g(X)]^2|X\}$$
$$E\{[Y-\mu(X)]^2\}\leqslant E\{[Y-g(X)]^2\}$$

最后这个性质在论述预测或预报问题时非常有用。第一个不等式是说，如果我们用以 X 为条件的均方预测误差作为预测不准确的一种度量，则条件均值比任何其他的 X 的函数作为 Y 的预测值都要好。条件均值还使得无条件均方预测误差最小化。

条件方差

给定随机变量 X 和 Y，以 $X=x$ 为条件的 Y 的方差就是在给定 $X=x$ 下相应的 Y 的条件分布的方差：$E\{[Y-E(Y|x)]^2|x\}$。公式

$$\mathrm{Var}(Y|X=x)=E(Y^2|x)-[E(Y|x)]^2$$

常用于计算。我们需要计算一个条件方差的机会并不多，但我们还必须对它做些假定，并对回归分析中的一些专题计算条件方差。

作为一个例子，令 $Y=SAVING$ 和 $X=INCOME$（每年都对全部家庭这个总体测量这两个变量）。假定 $\mathrm{Var}(SAVING/INCOME)=400+0.25INCOME$，这是说，随着收入增加，储蓄水平的方差也在增加。重要的是要看到 $SAVING$ 的方差和 $INCOME$ 之间的关系完全有别于 $SAVING$ 的期望值和 $INCOME$ 之间的关系。

我们叙述一个关于条件方差的有用性质。

性质 CV.1

如果 X 和 Y 相互独立，则

$$\mathrm{Var}(Y|X)=\mathrm{Var}(Y)$$

因为在给定 X 下 Y 的分布与 X 无关，而 $\mathrm{Var}(Y|X)$ 只不过是这个分布的一种特性，所以这个性质是相当明确的。

B.5 正态及其有关分布

正态分布

正态分布和由它衍生出来的分布是统计学和计量经济学中最广泛使用的分布。假定在总体上定义的随机变量是正态分布，将使概率计算得以简化。此外，为了进行统计学和计量经济学中的推断，即使背后的总体不一定是正 689 态的，我们仍然在很大程度上要依靠正态及其有关的分布。我们必须推迟对细节的讨论，但应肯定这些分布一直在本书中反复出现。

正态随机变量是一个可以取任何值的连续随机变量。它的概率密度函数有一个我们熟悉的钟形形状，如图 B.7 所示。

在数学上，X 的 pdf 可写为

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\{-(x-\mu)^2/2\sigma^2\},\quad -\infty<x<\infty \tag{B.34}$$

式中，$\mu=\mathrm{E}(X)$ 和 $\sigma^2=\mathrm{Var}(X)$。我们说 X 有一个均值为 μ 和方差为 σ^2 的**正态分布**（normal distribution），写作 $X\sim\mathrm{Normal}(\mu,\sigma^2)$。因正态分布对称于 μ，故 μ 也是 X 的中位数。有时又把正态分布叫做高斯分布，以纪念著名统计学家 C.F. 高斯（Gauss）。

一些随机变量粗略地看似乎遵循正态分布。人类的身高和体重、考试得分以及某县失业率，大体上都有类似于图 B.7 形状的 pdf。另一些分布如收

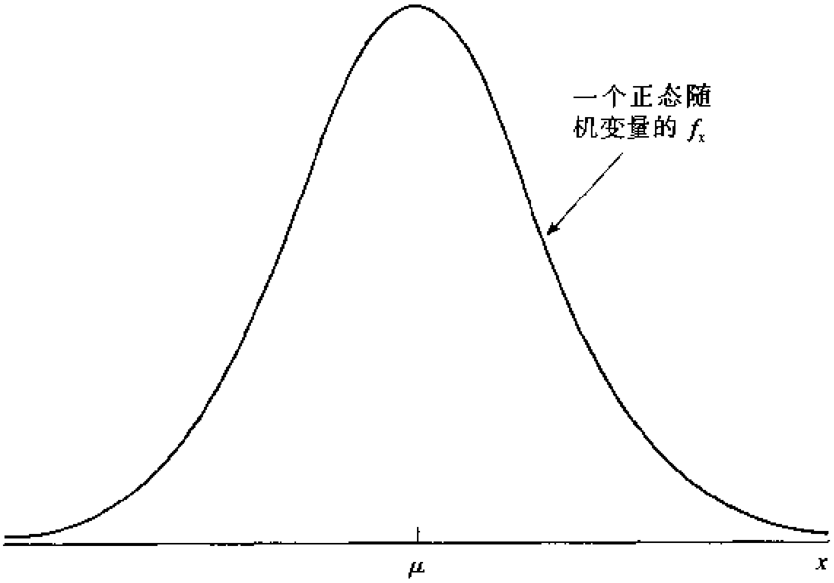


图 B.7 正态概率密度函数的一般形状

入分布，则不像正态密度函数那样分布。在大多数国家里，收入都不对称于任何数值而分布；分布是朝上端偏倚的。有时一个变量可通过变换而获得正态性。一个常见的变换是取自然对数，这对取正值的随机变量来说是有意义的。如果 X 是正的随机变量，诸如收入，而 $Y=\log(X)$ 是一个正态分布，就说 X 有一个**对数正态**（lognormal）分布。人们发现，对数正态分布颇适合于许多国家的收入分布。另一些变量，诸如商品的价格，看来也适合于描述为对数正态分布。

标准正态分布

690 正态分布的一种特殊情形是它的均值是零和方差（因而标准差）是 1。如果一个随机变量 Z 有一个 Normal（0，1）分布，我们就说它有一个**标准正态分布**（standard normal distribution），一个标准正态随机变量的 pdf 被记为 $\Phi(z)$；由式（B.34），它由 $\mu=0$ 和 $\sigma^2=1$ 的下式给出：

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\exp(-z^2/2),\quad -\infty<z<\infty \tag{B.35}$$

标准正态累积分布函数被记为 $\Phi(z)$，它被取为 Φ 之下 z 以左的面积；见图 B.8。请回忆 $\Phi(z)=P(Z\leqslant z)$；因 Z 是连续的，故也可写为 $\Phi(z)=P(Z<z)$。

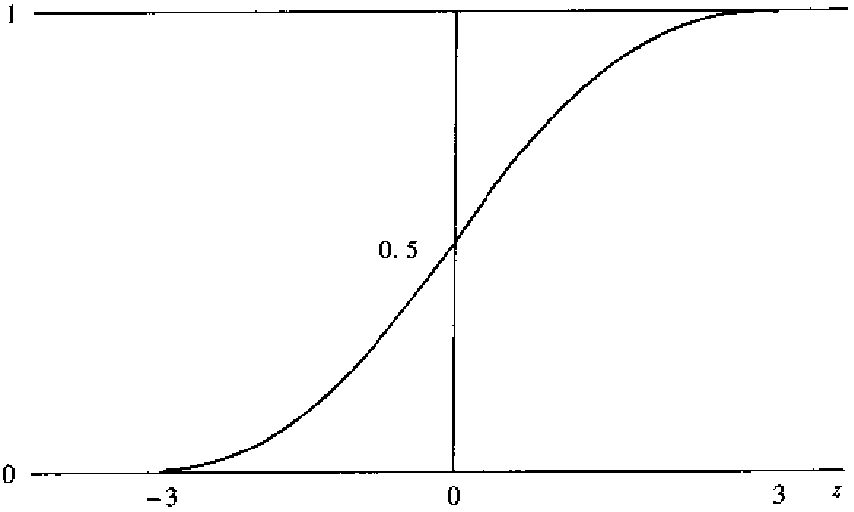


图 B.8　标准正态累积分布函数

691 没有可用来求 $\Phi(z)$ 的值的简单公式[因为 $\Phi(z)$ 是式(B.35)中的函数的积分，而这个积分没有一个封闭的形式]。然而 $\Phi(z)$ 的值是容易用表列出来的；表 G.1 给出了 z 从 -3.1 到 3.1 的 $\Phi(z)$ 值。对于 $z\leqslant -3.1$，$\Phi(z)$ 小于 0.001，而对于 $z\geqslant 3.1$，$\Phi(z)$ 大于 0.999。大多数统计学和计量经济学软件包都含有计算标准正态 cdf 值的简单命令，因此我们完全能避免使用排印好的表格而获得对应于任意 z 值的概率。

借助于概率的基本事实——特别是涉及 cdf 的性质(B.7)和(B.8)——我

们可以利用标准正态 cdf 计算涉及一个标准正态随机变量的任何事件的概率。最重要的公式是

$$P(Z>z)=1-\Phi(z) \tag{B.36}$$

$$P(Z<-z)=P(Z>z) \tag{B.37}$$

和

$$P(a\leqslant Z\leqslant b)=\Phi(b)-\Phi(a) \tag{B.38}$$

因为 Z 是连续的随机变量，所以不管不等式是否严格，这三个公式全都有效。比如 $P(Z>0.44)=1-0.67=0.33$，$P(Z<-0.92)=P(Z>0.92)=1-0.821=0.179$ 和 $P(1<Z\leqslant 0.5)=0.692-0.159=0.533$。

另一个有用的表达式是，对任何 $c>0$，有

$$\begin{aligned}P(|Z|>c)&=P(Z>c)+P(Z<-c)\\&=2\cdot P(Z>c)=2[1-\Phi(c)]\end{aligned} \tag{B.39}$$

就是说，Z 的绝对值大于某正常数 c 的概率就是概率 $P(Z>c)$ 的 2 倍；这反映了标准正态分布的对称性。

在大多数应用中，我们都从一个正态分布的随机变量 $X\sim \text{Normal}(\mu,\sigma^2)$ 开始，其中 μ 异于零且 $\sigma^2\neq 1$，然后利用以下性质可将其转变到一个标准正态分布。

性质 NORMAL.1

如果 $X\sim\text{Normal}(\mu,\sigma^2)$，则

$$(X-\mu)/\sigma\sim\text{Normal}(0,1)$$

性质 NORMAL.1 表明怎样把任一个正态随机变量转换为标准正态。例如，$X\sim\text{Normal}(3,4)$ 而我们要计算 $P(X\leqslant 1)$。我们总是把 X 规范化到一个标准正态变量：

$$\begin{aligned}P(X\leqslant 1)&=P(X-3\leqslant 1-3)=P\left(\frac{X-3}{2}\leqslant -1\right)\\&=P(Z\leqslant -1)=\Phi(-1)=0.159\end{aligned}$$

692

例 B.6　正态随机变量的概率

首先让我们计算当 $X\sim\text{Normal}(4,9)$ 时的 $P(2<X\leqslant 6)$（因为 X 是连续随机变量，所以用 $<$ 或 $\leqslant$ 都无关重要）。现在

$$\begin{aligned}P(2<X\leqslant 6)&=P\left(\frac{2-4}{3}<\frac{X-4}{3}\leqslant\frac{6-4}{3}\right)=P(-2/3<Z\leqslant 2/3)\\&=\Phi(0.67)-\Phi(-0.67)=0.749-0.251=0.498\end{aligned}$$

再让我们来计算 $P(|X|>2)$：

$$\begin{aligned}P(|X|>2)&=P(X>2)+P(X<-2)=2\cdot P(X>2)\\&=2\cdot P\left(\frac{X-4}{3}>\frac{2-4}{3}\right)=2\cdot P(Z>-0.67)\\&=2[1-\Phi(-0.67)]=0.772\end{aligned}$$

正态分布的其他性质

在结束本小节之前，我们列出以后会用得到的关于正态分布的其他若干性质。

性质 NORMAL.2

如果 $X\sim\text{Normal}\ (\mu,\ \sigma^2)$，则

$$aX+b\sim\text{Normal}\ (a\mu+b,\ a^2\sigma^2)$$

例如，若 $X\sim\text{Normal}\ (1,\ 9)$，则 $Y=2X+3$ 也是正态分布的，其均值为$2\text{E}(X)+3=5$，方差为 $2^2\cdot 9=36$，而 $\text{sd}(Y)=2\text{sd}(X)=2\cdot 3=6$。

前面我们讨论过，一般而言，零相关是怎样地不同于独立性的。而对于正态分布的随机变量而言，零相关却是独立性的充分条件。

性质 NORMAL.3

如果 X 和 Y 联合地正态分布，则它们是独立的，当且仅当 $\text{Cov}\ (X,\ Y)=0$。

性质 NORMAL.4

独立、同分布正态随机变量的任意线性组合是正态分布的。

例如 X_i（$i=1,\ 2,\ 3$）是遵从 $\text{Normal}\ (\mu,\ \sigma^2)$ 分布的独立随机变量。定义 $W=X_1+2X_2-3X_3$，则 W 是正态分布的；我们只须求出它的均值和方差。就是

$$\text{E}(W)=\text{E}(X_1)+2\text{E}(X_2)-3\text{E}(X_3)=\mu+2\mu-3\mu=0$$

并且

$$\text{Var}(W)=\text{Var}(X_1)+4\text{Var}(X_2)+9\text{Var}(X_3)=14\sigma^2$$

693 性质 NORMAL.4 还意味着独立、正态分布的随机变量的平均是一个正态分布变量。如果 $Y_1, Y_2, \cdots, Y_n$ 为独立随机变量，且每一变量都遵从 $\text{Normal}\ (\mu,\ \sigma^2)$ 而分布，则

$$\bar{Y}\sim\text{Normal}(\mu,\sigma^2/n) \tag{B.40}$$

这一结果在对正态总体中的均值作统计推断时起关键作用。

χ 平方分布

χ 平方分布可直接从独立标准正态随机变量推出。令 $Z_i(i=1,2,\cdots,n)$ 为独立随机变量，各遵循标准正态分布。定义一个新随机变量为诸 Z_i 平方之和：

$$X=\sum_{i=1}^{n}Z_i^2 \tag{B.41}$$

于是，称 X 为有 n 个自由度（或简记 df）的 χ 平方分布。记为 $X\sim\chi^2_n$。χ

平方中的 df 对应于总和（B.41）中的项数。自由度的概念在我们的统计和计量经济分析中扮演重要角色。

带有不同自由度的 χ 平方 pdf 的图形，由图 B.9 给出；我们不需要这个 pdf 的公式，所以不把它在这里重写出来。由方程（B.41）明显看到，χ 平方随机变量总是非负的，而且不像正态分布那样，χ 平方分布不对称于任何点。可以证明，如果 $X \sim \chi_n^2$，则 X 的期望值是 n［式（B.41）中的项数］，而 X 的方差是 $2n$。

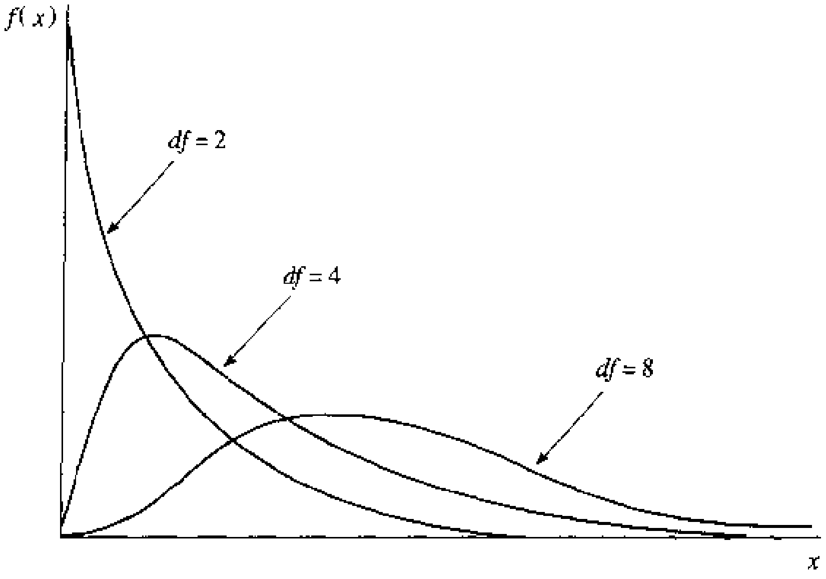


图 B.9　有各种自由度的 χ 平方分布

t 分布

t 分布是经典统计学和多元回归分析中的负重者；可以从一个标准正态和一个 χ 平方分布得到它。

设 Z 有标准的正态分布，而 X 有自由度为 n 的 χ 平方分布。于是，随机变量

$$T = \frac{Z}{\sqrt{X/n}} \tag{B.42}$$

便有自由度为 n 的 t 分布，我们把它记为 $T \sim t_n$。t 分布是从 χ 平方随机变量那里得到它在式（B.42）的分母中的自由度的。

t 分布的 pdf 有一个类似于标准正态分布那样的形状，只是它更散开一些，因而尾端有较大的面积。t 分布随机变量的期望值为零（严格说，期望
694 值仅当 $n>1$ 时存在），方差为 $n/(n-2)$（这里要求 $n>2$；当 $n \leqslant 2$ 时，分布没有散开，方差将不存在）。图 B.10 对不同的自由度描绘了 t 分布的 pdf。随着自由度变大，t 分布接近于标准正态分布。

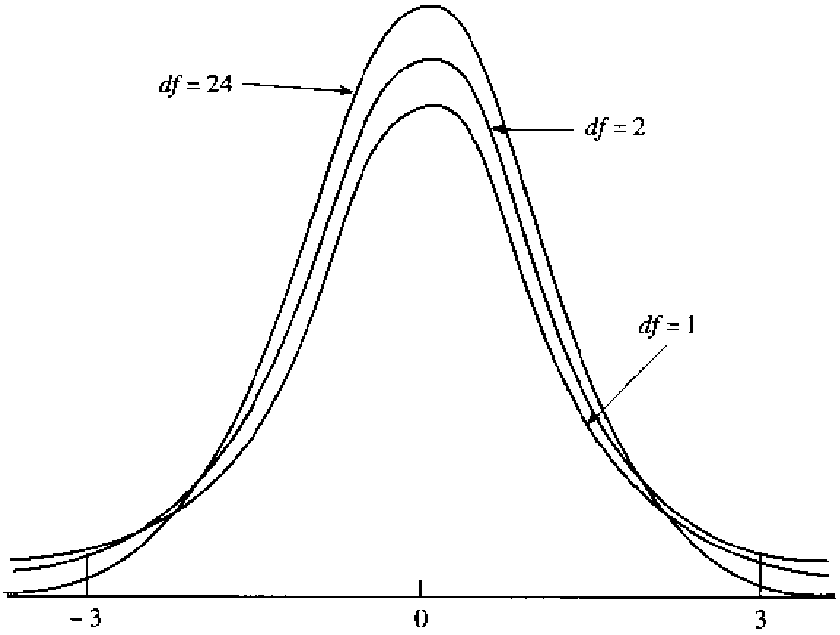


图 B.10　各种自由度的 t 分布

F 分布

统计学和计量经济学的另一个重要分布是 F 分布。特别是在多元回归分析领域中，要用 F 分布去检验假设。

为了定义 F 随机变量，令 $X_1 \sim \chi^2_{k1}$ 和 $X_2 \sim \chi^2_{k2}$，并假定 X_1 和 X_2 独立，则随机变量

$$F = \frac{(X_1/k_2)}{(X_2/k_2)} \tag{B.43}$$

有一个自由度为（k_1，k_2）的 F 分布。我们把它记为 $F \sim F_{k_1k_2}$。图 B.11 对

695

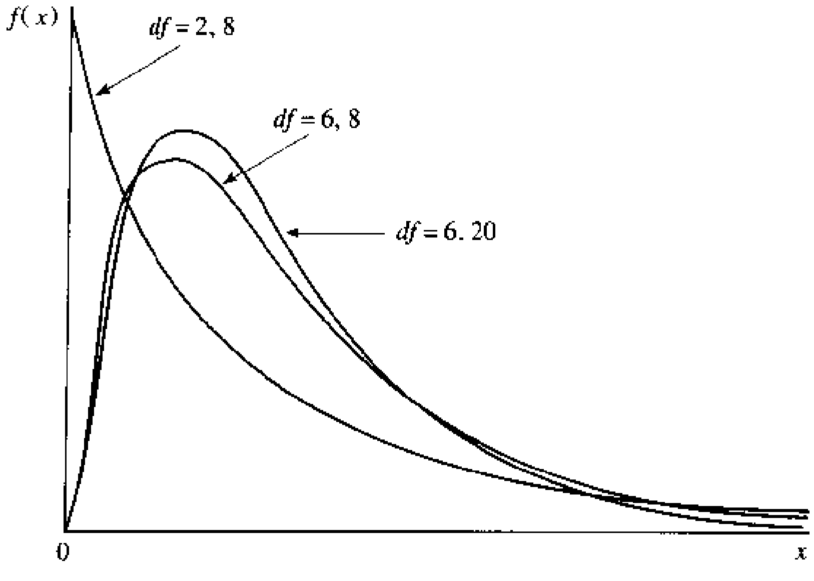


图 B.11　各种自由度 k_1 和 k_2 的 $F_{k_1k_2}$ 分布

不同的自由度描绘了 F 分布的 pdf。

自由度的顺序是关系重大的。整数 k_1 由于联系到分子中的 χ 平方变量而常称为分子自由度。类似地，整数 k_2 由于联系到分母中的 χ 平方变量而称为分母自由度。这似乎有点玄妙，因为式（B.43）又可写为（X_1k_2）/（X_2k_1），这样一来，k_1 反而出现在分母中了。请仅记住，分子 *df* 是指式（B.43）分子中的 χ 平方变量的 *df*，并类似地解释分母 *df*。

▶ 小结

在本附录中，我们复习了计量经济学中所需要的概率概念。你应该从概率统计入门课程中熟悉其中的大部分。某些较高深的问题，诸如条件期望，还不需要现在就掌握，可以等到这些概念出现在本书第Ⅰ部分讨论回归分析的时候。

在入门的统计学课程中，注意点是放在对一些特殊分布计算其均值、方差、协方差等等上面。在本书第Ⅰ部分中，不需要这样的计算；我们主要依靠本附录中所陈述的关于期望值、方差等等的性质。 696

关键术语

贝努里（或二值）随机变量
二项分布
χ 平方分布
条件分布
条件期望
连续随机变量
相关系数
协方差
累积分布函数（cdf）
自由度
离散随机变量
期望值
实验
F 分布
独立随机变量
联合分布
重期望律
中位数
正态分布
两两不相关随机变量
概率密度函数（pdf）
随机变量
标准差
标准正态分布
标准化随机变量
t 分布
方差

习　题

697 **B.1** 假定一名中学生准备参加 *SAT* 考试。试解释他或她的最终 *SAT* 分数适合于看做一个随机变量。

B.2 令 X 为一个遵循 Normal（5，4）分布的随机变量。求以下事件的概率：

(i) $P(X \leqslant 6)$

(ii) $P(X > 4)$

(iii) $P(|X-5| > 1)$

B.3 某些互助基金的运作年复一年地胜过市场的运作（意为持有互助基金的股份比持有诸如 S&P500 的组合证券有更高的回报），这一事实引起了许多讨论。为具体起见，考虑 10 年的期间，并令总体为华尔街时报在 1995 年 6 月 1 日报导的 4 170 个互助基金。我们说相对于市场的运作是随机的，意为每个基金任一年都有 50－50 的（百分数）机会运作得比市场更好，而且各年的运作是独立的。

(i) 如果相对于市场的运作是真正随机的，那么任一特定基金在整个 10 年里都运作得比市场好的概率是什么？

(ii) 4 170 个基金中至少有一个基金在整个 10 年里都运作得比市场好的概率又是什么？从你的答案中你发现了什么？

(iii) 如果你有一个能计算二项式概率的统计包，求至少有 5 个基金在整个 10 年期间都运作得比市场要好的概率。

B.4 对美国的一个随机抽取的县，令 X 代表 65 岁以上的就业者在成年人中占的比例，或者说老年就业率。因此 X 的值限制在 0～1 之间。假定 X 的累积分布函数由 $F(x) = 3x^2 - 2x^3$（$0 \leqslant x \leqslant 1$）给出。求老年就业率至少是 0.6（即 60%）的概率。

B.5 1995 年，刚好在挑选 O.J. 辛普森谋杀审判案的陪审员之前，一次民意测验发现，约有 20% 的成年人认为辛普森是无罪的（在该案的许多人证物证已向公众透露之后）。暂且忽略这个 20% 是根据总体的一个子样本估计出来的事实；为了说明概念，不妨把它看做在挑选陪审员之前相信辛普森无罪的人占的真正百分数。假定从总体中随机并且独立地挑选（尽管事实并非如此）12 个陪审员。

(i) 求陪审团中至少有 1 人在陪审员挑选之前认为辛普森无罪的概率。[提示：定义这个 Binomial（12，20）随机变量 X 为相信辛普森无罪的陪审员人数。]

(ii) 求陪审团中至少有 2 人相信辛普森无罪的概率。[提示：$P(X \geqslant 2) = 1 - P(X \leqslant 1)$，而 $P(X \leqslant 1) = P(X=0) + P(X=1)$]

698 **B.6** （需要用微积分）令 X 为美国某州被判有汽车盗窃罪的人要在监

狱中服刑的年数。假定 X 的 pdf 为

$$f(x) = (1/9)x^2, \quad 0 < x < 3$$

用积分法求期望的监禁年数。

B.7 如果一名篮球运动员的罚球命中率是 74%，那么，他或她在一次有 8 个罚球机会的比赛中，平均而言将会投进多少个球？

B.8 假定一名大学生正在修读三门课：一门 2 学分、一门 3 学分和一门 4 学分。2 学分课的期望成绩是 3.5，而 3 学分和 4 学分课的期望成绩都是 3.0。问该学期他的总学分成绩的期望值是什么？［提示：注意每门课的成绩要用它在总学分中所占的份额来加权。］

B.9 令 X 为美国大学教授以千美元计的年薪。假定平均年薪是 52.3，年薪的标准差是 14.6。求以美元计的年薪的均值和标准差。

B.10 假定在一所大型的大学里，大学学分平均成绩 GPA 和中学业绩得分 SAT 的关系由条件期望 $\mathrm{E}(GPA/SAT) = 0.70 + 0.002SAT$ 给出。

(i) 求当 $SAT = 800$ 时的期望 GPA。求 $\mathrm{E}(GPA \mid SAT = 1\ 400)$，评论这一差异。

(ii) 如果该大学的平均 SAT 是 1 100，那么平均 GPA 是多少？［提示：利用性质 CE.4。］

附录 C　数理统计基础

C.1　总体、参数与随机抽样

699　**统计推断**（statistical inference）指利用来自某**总体**（population）的一个样本，从中获知该总体的某些情况。所谓总体，乃指任何良好定义的一组对象，这些对象可以是个人、厂商、城市或其他别的什么，所谓“获知”，可以有多种含义，但大致归类为**估计**（estimation）和**假设检验**（hypothesis testing）两个范畴。

举一两个例子也许能帮助你理解这些词句。劳工经济学家想获知美国全体成年劳工的教育回报，问再多受一年教育，工资平均会增加百分之几？要获得美国全体劳工的工资和教育信息是不现实和不经济的，但我们可以获得总体的一个子集的数据。利用收集到的这些数据，一位劳工经济学家也许能报导他或她对再受一年教育的回报的最好估计为 7.5%。这就是**点估计**（point estimate）的一个例子。或者，她或他报告一个区域，比方说“教育的回报在 5.6%～9.4%之间”，这是**区间估计**（interval estimate）的一个例子。

城市经济学家也许想知道邻居犯罪监视计划是否与低犯罪率有关。经过

在取自总体的一个样本中比较了安排和不安排监视计划的邻居犯罪率，他或她可以作出两结论之一：邻居监视计划对犯罪率确有影响，或者没有影响。这个例子就落在假设检验的标识下。

统计推断的第一步是要明确所关注的总体。这也许是不言自明的，但务必使之非常明确才行。一旦明确了总体是什么，就可对所关注的总体关系建立或设定一个模型。这个模型将涉及一些概率分布或概率分布的特征，而这又取决于一些未知参数。所谓参数就是决定变量之间的关系式的方向和强度的一些常数。在上述劳工经济学的例子中，所关注的参数是总体中的教育回报（率）。

抽 样

700 为了复习统计推断的概念，我们聚焦在最可能简单的一种格局上。令 Y 代表概率密度函数为 $f(y;\theta)$ 的总体的一个随机变量，其中 $f(y;\theta)$ 依赖于单一个参数 θ。假定除了 θ 值未知外，Y 的概率密度函数（pdf）是已知的。不同的 θ 值将意味着不同的概率分布，因此我们对 θ 值感兴趣。如果我们能得到该总体的某种样本，我们就能获知 θ 的某些情况。最容易处理的抽样方案是随机抽样。

随机抽样

如果 $Y_1, Y_2, \cdots, Y_n$ 是有同一概率密度函数 $f(y;\theta)$ 的独立随机变量，我们就说 $\{Y_1, Y_2, \cdots, Y_n\}$ 是一个来自 $f(y;\theta)$ 的**随机样本**（random sample）[或者说来自由 $f(y;\theta)$ 来代表的总体的随机样本]。

当 $\{Y_1, \cdots, Y_n\}$ 是来自密度 $f(y;\theta)$ 的一个随机样本时，我们又说诸 Y_i 是取自 $f(y;\theta)$ 的**独立、同分布**（independent, identically distributed, $i.i.d.$）样本。在一些情况中，我们将不需要完全设定这个共同的分布是什么。

在随机抽样的定义中，$Y_1, Y_2, \cdots, Y_n$ 的随机性质反映了这样的事实：在抽样实际完成之前，许许多多的不同结果都是可能的。例如，我们获取了 $n=100$ 个美国家庭的家庭收入，那么我们观测到的收入将随每个不同的100户样本而不同。一旦得到了一个样本，我们就得到一个数集，比方说 $\{Y_1, Y_2, \cdots, Y_n\}$，这就是我们要加以研究的数据。把这个样本假定为来自一个随机抽样方案是否得当，要求我们具有关于实际抽样过程的知识。

取自贝努里分布的随机样本常被用来说明一些统计概念，并且这种样本也会出现在实证应用之中。如果 $Y_1, Y_2, \cdots, Y_n$ 是独立随机变量，并且每一个都是按照Bernoulli(θ)分布的，使得 $P(Y_i=1)=\theta$ 而 $P(Y_i=0)=1-\theta$，那么 $\{Y_1, Y_2, \cdots, Y_n\}$ 就构成一个来自Bernoulli(θ)分布的随机样本。作为一个说明例子，考虑附录B讨论过的飞机订座例子。令 Y_i 代表订了座位的乘客 i 到时是否出现的随机变量：如果乘客 i 出现则 $Y_i=1$；否则 $Y_i=0$。在这里，θ 是从所有订了座位的人的总体中随机抽取的一人到时来乘坐飞机的概率。

在许多的其他应用中，可以假定随机样本是取自一个正态分布（的总体）的。如果$\{Y_1, Y_2, \cdots, Y_n\}$是一个来自 Normal$(\mu, \sigma^2)$总体的随机样本，则该总体由两个参数即均值$\mu$和方差$\sigma^2$来划定。人们主要关注的常常是$\mu$，不过$\sigma^2$本身也是重要的。要对$\mu$作出推断常常需要关于$\sigma^2$的知识。

C.2 估计量的有限样本性质

在本节中我们研究所谓的估计量的有限样本性质。“有限样本”（finite sample）一词来自如下事实：所讨论的性质对任何大小的样本都成立，且无论样本有多大或多小。有时又把这些性质叫做小样本性质。在 C.3 节中，我 701 们将讨论随着样本大小的无限增大时，估计量将要表现出来的“渐近性质”（asymptotic properties）。

估计量与估计值

为了研究估计量的性质，我们必须定义什么是估计量。给定一个来自依赖于某未知参数θ的总体的随机样本$\{Y_1, Y_2, \cdots, Y_n\}$，$\theta$的一个**估计量**（estimator）就是赋予样本的每一可能结果一个θ值的法则。这个法则是抽样进行之前就已制定好的；特别是不管实际上得到的是什么数据，这个法则都不变。

作为估计量的一个例子，令$\{Y_1, Y_2, \cdots, Y_n\}$为取自均值为$\mu$的总体的一个随机样本。$\mu$的一个很自然的估计量将是这个随机样本的均值：

$$\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i \tag{C.1}$$

我们把$\bar{Y}$叫做**样本均值**（sample average），但是，它不同于我们在附录 A 中作为一个描述统计量而定义的一个数集的样本均值。现在要把$\bar{Y}$看做一个估计量。给定随机变量$Y_1, \cdots, Y_n$的任何一种结果，我们都用同样的法则去估计μ：就是取它们的平均。对于实际结果$\{y_1, \cdots, y_n\}$，**估计值**（estimate）就是该样本的均值：$\bar{y} = (y_1 + y_2 + K + y_n)/n$。

例 C.1　城市失业率

假使我们得到美国 10 个城市失业率的如下样本：

城市	失业率
1	5.1
2	6.4

3	9.2
4	4.1
5	7.5
6	8.3
7	2.6
8	3.5
9	5.8
10	7.5

702 我们对美国平均城市失业率的估计值是 $\bar{y}=6.0$。一般地说，每个样本都有一个不同的估计值，但是求估计值的法则是一样的，不管在样本中出现的是哪些城市，也不管样本中有多少个城市。

更一般地说，参数 θ 的一个估计量 W 可表示为一个抽象的数学公式：

$$W=h(Y_1,Y_2,\cdots,Y_n) \tag{C.2}$$

式中，h 为随机变量 $Y_1,Y_2,\cdots,Y_n$ 的某个已知函数。如同样本均值的特殊情形那样，W 因为有赖于随机样本所以是一个随机变量：W 随着从总体中抽到不同的随机样本而可能改变。当我们把一个特定的数集，比方说 $\{y_1, y_2, \cdots, y_n\}$ 代入函数 h 中时，得到 θ 的一个估计值，记为 $w=h(y_1,\cdots,y_n)$。有时把 W 叫做点估计量而把 w 叫做点估计值，以区别于 C.4 节将要讲的区间估计量和估计值。

为了评价不同的估计方法，我们研究随机变量 W 的概率分布的各种性质。一个估计量的分布常称做**抽样分布**（sampling distribution）。原因是这个分布描述了缘于不同随机样本的各种 W 结果的似然性。因为如何组合数据以估计参数的法则可以有无限之多，我们需要一些有意义的准则用以挑选估计量，或者至少能淘汰一些估计量。于是，我们必须告别描述统计量这一领域，不再仅为总结一组数据而计算诸如样本均值一类的东西。在数理统计学中，我们研究的是估计量的抽样分布。

无偏性

原理上，给定 Y_i 的概率分布和函数 h，我们能求出 W 的整个抽样分布。通常，在评价 W 作为 θ 的一个估计量时，集中考虑 W 分布的少数几个特征是比较容易的。一个估计量的第一重要性质就是关于它的期望值。

无偏估计

θ 的估计量是无偏的，如果对一切可能的 θ 值，都有

$$\mathrm{E}(W)=\theta \tag{C.3}$$

如果一个估计量是无偏的，它的概率分布就有一个等于它所估计的参数的期望值。**无偏性**（unbiasedness）并不是说我们用某一特定样本得到的估计值等于 θ，或者甚至说很接近 θ，而是说如果我们能够从总体中抽取关于 Y 的无限多个样本的话，每次都计算一个估计值，那么对所有的随机样本把这些估计值平均起来，我们就会得到 θ。这个想像中的实验是抽象的，因为在大多数应用中，我们使用的仅仅是一个随机样本。

对于一个不是无偏的估计量，我们定义它的偏误如下。

一个估计量的偏误

如果 W 是 θ 的一个估计量，则它的**偏误**（bias）被定义为

$$\text{Bias}(W) \equiv \text{E}(W) - \theta \qquad \text{(C.4)}$$

图 C.1 展示两个估计量，第一个是无偏的，而第二个有一正的偏误。

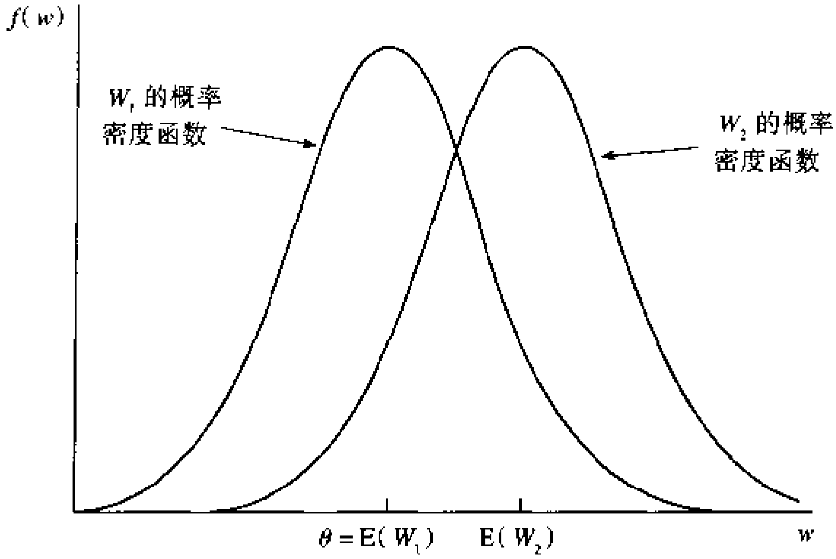


图 C.1 一个无偏估计量 W_1 和一个有正偏误的估计量 W_2

一个估计量的无偏性和可能偏误的大小有赖于 Y 的分布和函数 h。通常，Y 的分布不是我们所能控制的（虽然我们常常为这个分布选择一个模型）：它也许由自然规律或社会力量来决定。但法则 h 的选择则操纵在我们手中，如果我们想要一个无偏估计量，我们就要对 h 作相应的选择。

可以证明，一些估计量在很一般的情形下是无偏的。现在我们来证明，样本均值 $\bar{Y}$ 是总体均值 μ 的一个无偏估计量，不管这个背后的总体是怎样分布的。利用我们在 B.3 节中讨论的期望值性质 E.1 和 E.2，有

$$\begin{aligned}\text{E}(\bar{Y}) &= \text{E}\left[(1/n)\sum_{i=1}^{n} Y_i\right] = (1/n)\text{E}\left(\sum_{i=1}^{n} Y_i\right) \\ &= (1/n)\left[\sum_{i=1}^{n}\text{E}(Y_i)\right] \\ &= (1/n)\left(\sum_{i=1}^{n}\mu\right) = (1/n)(n\mu) = \mu\end{aligned}$$

为了做假设检验，我们还有必要从均值为 μ 的总体中估计方差 σ^2。令
704 $\{Y_1,\cdots,Y_n\}$ 为取自 $E(Y)=\mu$ 和 $\mathrm{Var}(Y)=\sigma^2$ 的总体的随机样本，定义估计
量为

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 \tag{C.5}$$

它常被称为**样本方差**（sample variance）。可以证明，S^2 是 σ^2 的无偏估计量：$E(S^2)=\sigma^2$。用 $n-1$ 而不用 n 为除数是因为式中用的是均值 μ 的估计量而不是它的真值。如果 μ 已知，那么 σ^2 的一个无偏估计量将是 $n^{-1}\sum_{i=1}^{n}(Y_i-\mu)^2$。然而实践中很少知道 μ。

虽然无偏性作为估计量的一个性质颇具魅力——它的反义词"偏误"确实有些消极方面的含义——但它也不是没有自身的问题的。无偏性的一个弱点是，一些合理的甚至是相当好的估计量却不是无偏的。我们很快要看到这方面的一个例子。

无偏性的另一重要弱点是，实际上存在很不好的无偏估计量。考虑我们在估计一个总体的均值 μ。假定我们不用样本均值 $\bar{Y}$ 去估计 μ，而是在收集一个大小为 n 的样本后，只保留第一个观测值并删除掉其余的，然后用 $W\equiv Y_1$ 作为 μ 的估计量，因为 $E(Y_1)=\mu$，所以这个估计量也是无偏的。很可能你意识到了忽略除第一个以外的所有观测值，不是一个明智的估计方法：它把样本中的大部分信息都丢掉了。例如，取 $n=100$，我们有随机变量 Y 的 100 个观测结果，可是我们只用其中的第一个去估计 $E(Y)$。

估计量的抽样方差

上小节末的例子表明，我们还需要用更多的准则来评价一个估计量的好
坏。无偏性仅保证估计量的概率分布有一个等于它要估计的参数的均值。这
自然是好的，但我们还须知道这个估计量的分布究竟有多分散。一个估计量
可以在平均意义下等于 θ，但它仍然会以大概率偏离到很远处。在图 C.2
705 中，W_1 和 W_2 都是 θ 的无偏估计量，但 W_1 的分布更紧密地集中在 θ 的周
围：W_1 对 θ 的距离大于任何给定值的概率都要小于 W_2 对 θ 的距离大于同
样给定值的概率。用 W_1 作为估计量就意味着，我们将有较小的可能性获得
一个产生远离 θ 的估计值的随机样本。

为了概括图 C.2 所展示的情况，我们要用到一个估计量的方差（或标准差）。记得这是分布中的散度的一个度量。估计量的方差常叫做**抽样方差**（sampling variance），因为这个方差是和抽样分布联系起来的。要记住，抽样方差不是一个随机变量；它是一个常数，也许是未知的。

我们现在来求用以估计总体均值 μ 的样本均值的方差：

$$\begin{aligned}\text{Var}(\bar{Y}) &= \text{Var}\Big[(1/n)\sum_{i=1}^{n} Y_i\Big] = (1/n^2)\text{Var}\Big(\sum_{i=1}^{n} Y_i\Big)\\ &= (1/n^2)\Big[\sum_{i=1}^{n}\text{Var}(Y_i)\Big] = (1/n^2)\Big(\sum_{i=1}^{n}\sigma^2\Big)\\ &= (1/n^2)(n\sigma^2) = \sigma^2/n \qquad (\text{C}.6)\end{aligned}$$

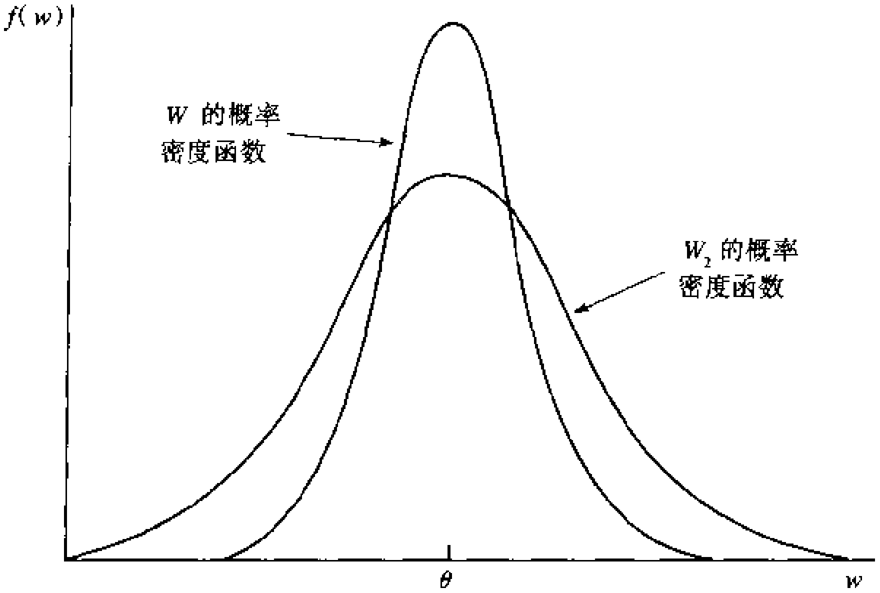


图 C.2　θ 的两个无偏估计量的抽样分布

706 注意我们是怎样地使用了 B.3 节和 B.4 节（VAR.2 和 VAR.4）以及 Y_i 的独立性的。概括起来：如果 $\{Y_i: i=1,2,\cdots,n\}$ 是取自均值为 μ 和方差为 σ^2 的总体的一个随机样本，那么 $\bar{Y}$ 有和总体一样的均值，但它的方差等于总体方差 σ^2 除以样本大小 n。

$\text{Var}(\bar{Y}) = \sigma^2/n$ 的一个重要含义是，增大样本大小可以使它非常接近零。这是一个合理估计量的关键性特征，我们将在 C.3 节再回到这个问题。

如图 C.2 所提示的，在无偏一类估计量中，选取有最小方差的估计量，就使我们能淘汰某些估计量而不予考虑。对一个来自均值为 μ 和方差为 σ^2 的总体的随机样本，我们知道 $\bar{Y}$ 是无偏的，并且 $\text{Var}(\bar{Y}) = \sigma^2/n$。那么估计量 Y_1 这个仅仅是第一次抽取的观测值又怎样呢？既然 Y_1 是来自总体的一个随机抽取，就应有 $\text{Var}(Y_1) = \sigma^2$。因此，即令样本较小，$\text{Var}(Y_1)$ 和 $\text{Var}(\bar{Y})$ 之间的差异也可能大。如果 $n = 10$，那么 $\text{Var}(Y_1)$ 就 10 倍于 $\text{Var}(\bar{Y})\sigma^2/10$。这就给了我们排除掉 Y_1 作为 μ 的估计量的一个规范方法。

为了强调这一点，表 C.1 给出了一个小型仿真研究的结果。利用统计软件包 Stata，产生了样本大小为 10 的来自 $\mu = 2$ 和 $\sigma^2 = 1$ 的总体的 20 个随机样本；我们的兴趣在于估计 μ。对 20 个随机样本中的每一个，我们都计算两个估计值 y_1 和 $\bar{y}$，这些值列在表 C.1 中。从表中看到，y_1 的值远比 $\bar{y}$ 的分散：y_1 从 -0.64 变到 4.27，而 $\bar{y}$ 只从 1.16 变到 2.58。此外，20 个样本中有 16 个样本的 $\bar{y}$ 比 y_1 更接近于 $\mu = 2$。y_1 对所有的仿真求平均约为

1.89，而 $\bar{y}$ 的同样平均为 1.96。这两个平均数都接近于 2，说明了两个估计量的无偏性。（通过做更多的重复，还能使这些均值更接近 2。）但是只比较这些随机抽样的平均值，会抹杀样本均值 $\bar{Y}$ 作为 μ 的估计量远远胜过 Y_1 这一事实。

表 C.1　对 $\mu=2$ 的一个 Normal(μ,1) 分布的估计量仿真

重复	y_1	$\bar{y}$
1	−0.64	1.98
2	1.06	1.43
3	4.27	1.65
4	1.03	1.88
5	3.16	2.34
6	2.77	2.58
7	1.68	1.58
8	2.98	2.23
9	2.25	1.96
10	2.04	2.11
11	0.95	2.15
12	1.36	1.93
13	2.62	2.02
14	2.97	2.10
15	1.93	2.18
16	1.14	2.10
17	2.08	1.94
18	1.52	2.21
19	1.33	1.16
20	1.21	1.75

707

有效性

比较前一小节中 $\bar{Y}$ 和 Y_1 的方差是比较不同的无偏估计量的一般途径的一个例子。

相对有效性

假如 W_1 和 W_2 是 θ 的两个无偏估计量，那么，如果对所有的 θ 都有 $\mathrm{Var}(W_1) \leqslant \mathrm{Var}(W_2)$ 且至少对 θ 的一个值严格不等式成立，则称 W_1 比 W_2 更有效。

708 前面我们表明了，为估计总体均值 μ，只要 $n>1$，对任何 σ^2 值都有 $\mathrm{Var}(\bar{Y})<\mathrm{Var}(Y_1)$。但我们不能一味根据最小方差准则在无偏估计量之中进行选择。给定 θ 的两个无偏估计量，一个可能对一些 θ 值有较小方差，而另一个对另一些 θ 值有较小方差。

如果我们限于考虑某一类的估计量，我们就能证明样本均值有最小方差。习题 C.2 要求你证明，在所有无偏估计量，同时又是 $Y_1, Y_2, \cdots, Y_n$ 的线性函数中，$\bar{Y}$ 有最小方差。这里的假定是，诸 Y_i 有共同的均值和方差，并且它们两两不相关。

如果不限于考虑无偏估计量，那么比较方差大小就毫无意义。例如，在估计总体均值 μ 时，我们可以考虑一个微不足道的估计量：不管我们抽取到什么样本，这个估计量总取为零。自然而然，这个估计量的方差是零（因为对每个随机样本它都是同一个值）。但这个估计量的偏误是 $-\mu$，当 $|\mu|$ 很大时，它是一个很糟糕的估计量。

比较不一定是无偏的估计量的一个方法，是计算估计量的**均方差**（mean squared error，MSE）。如果 W 是 θ 的一个估计量，则它的 MSE 被定义为 $\mathrm{MSE}(W)=\mathrm{E}[(W-\theta)^2]$。MSE 度量着估计量 W 离开 θ 的平均距离。可以证明，$\mathrm{MSE}(W)=\mathrm{Var}(W)+[\mathrm{Bias}(W)]^2$。因此 $\mathrm{MSE}(W)$ 依赖于方差和偏误（如果有偏误的话）。这就使得我们能比较任何两个估计量，即使其中之一或两者都是有偏误的。

C.3　估计量的渐近或大样本性质

在 C.2 节中我们看到，作为总体均值 μ 的估计量 Y_1 尽管是无偏的，却是一个糟糕的估计量，它的方差可能比样本均值的方差大得多。Y_1 的一个明显特征是它对任何大小的样本都有同样的方差。看来，要求任何估计方法必须随样本的增大而改进是合理的。为了估计总体均值 μ，$\bar{Y}$ 的改进在于它的方差随 n 的变大而减少；而 Y_1 就没有这种改进。

研究估计量的**渐近**（asymptotic）或**大样本**（large sample）性质可以避免一些笨拙的估计量。此外，还能对不是无偏的和不容易求出其方差的估计量作些积极的考虑。

渐近分析涉及如何迫近一个估计量的抽样分布的特征。近似程度依赖于样本的大小。不幸的是，要说出多大的样本才能使渐近分析适当，我们必然感到受拘束；这要看背后的总体分布是什么而定。然而，我们已经知道，对小到 $n=20$ 的样本大小，大样本近似有时（“有时”为译者所加）也会运作

得不错。

一致性

估计量的第一个渐近性质告诉我们，随着样本大小无限地增加，用以估计某参数的估计量会离开该参数多远。

一致性

709 令 W_n 为基于大小为 n 的一个样本 $Y_1, Y_2, \cdots, Y_n$ 对 θ 的一个估计量。那么，如果随着 $n \to \infty$，对于每一个 $\varepsilon > 0$，都有

$$P(|W_n - \theta| \geqslant \varepsilon) \to 0 \tag{C.7}$$

就说 W_n 是 θ 的一个**一致（性）估计量**（consistent estimator）。如果 W_n 对 θ 不是一致（性）的，则说它是**非一致（性）的**（inconsistent）。

当 W_n 是一致的，我们也说 θ 是 W_n 的**概率极限**（probability limit），写作 $\text{plim}(W_n) = \theta$。

和无偏性不一样——无偏性是估计量在给定样本大小之下的一个特征——一致性描述着当样本大小变大时估计量的抽样分布的性态。为了强调这一点，我们在陈述上述定义时，就已对估计量加上样本大小 n 这个下标，并将在本节中始终保持这个惯常的做法。

方程（C.7）来看颇具技术性，因此建立一些基于基本概率的原理颇为困难。相反，要解释方程（C.7）则是容易的。它的意思是，W_n 的分布越来越集中于 θ，粗略地说，也就是对于越来越大的样本，W_n 离开 θ 很远的可能性越来越小。图 C.3 说明了这个趋势。

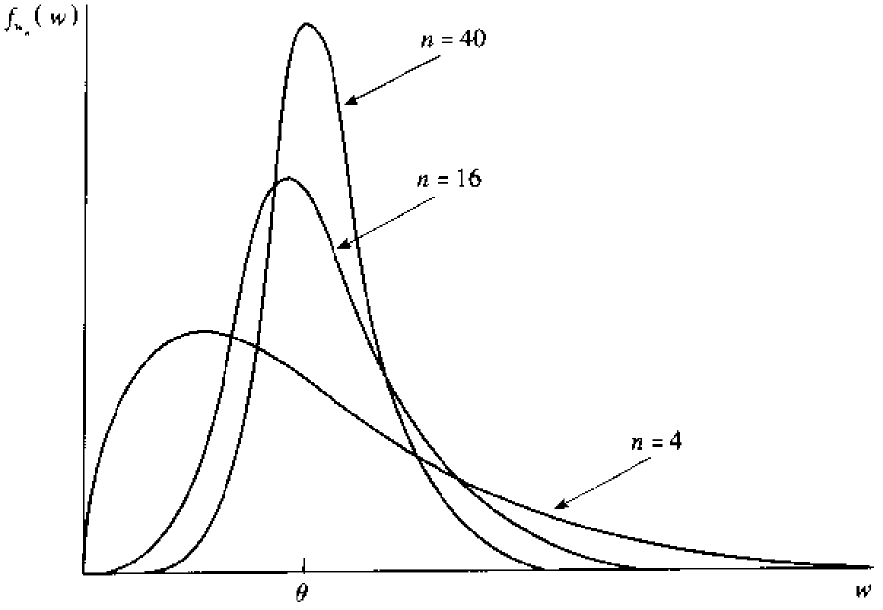


图 C.3　三种样本大小的一致估计量的抽样分布

710 如果一个估计量不是一致的，那么即使我们有无限多的数据，也无助于获知 θ。为了这个原因，一致性就成为统计学或计量经济学中对所用的估计量的一个起码要求。我们将遇到一些估计量在某些假定下是一致的，而当这些假定不成立时又是不一致的。当估计量是不一致的时，我们通常能求出它的概率极限，以便知道这些概率极限究竟离 θ 多远，这点是重要的。

如我们前面所指出的，无偏估计量不一定是一致的，但那些随样本大小增大而方差缩减到零的无偏估计量是一致的。这个结果可正式陈述如下：如果 W_n 是 θ 的无偏估计量，并且随着 $n \to \infty$ 而 Var（W_n）$\to 0$，则 plim（W_n）$= \theta$。大凡利用全部数据样本的无偏估计量通常都有一个随着样本大小增大而缩减到零的方差，所以是一致的。

一致估计量的一个很好的例子，是取自均值为 μ、方差为 σ^2 的总体的一个随机样本的均值。我们已经证明该样本均值对 μ 是无偏的。在方程（C.6）中，我们对任何一个样本大小 n 导出了 Var（Y_n）$= \sigma^2/n$。因此，随着 $n \to \infty$，有 Var（Y_n）$\to 0$。而且，$\bar{Y}_n$ 除了无偏外，还是 μ 的一致估计量。

$\bar{Y}_n$ 是 μ 的一致估计量，这一结论即使 Var（$\bar{Y}_n$）不存在也是成立的。这个经典性结果曾以**大数定律**（law of large numbers, LLN）为名。

大数定律

令 $Y_1, Y_2, \cdots, Y_n$ 是均值为 μ 的独立、同分布随机变量，于是

$$\text{plim}(\bar{Y}_n) = \mu \tag{C.8}$$

大数定律的意思是，如果我们对总体均值 μ 的估计感兴趣，那么通过选取一个足够大的样本，就能得到一个任意接近 μ 的数。这个奠基性的结果和 plim 的基本性质结合起来了，可以证明一些相当复杂的估计量是一致的。

性质 PLIM.1

令 θ 为某参数并定义一个新参数 $\gamma = g(\theta)$，$g(\theta)$ 代表某连续函数。假定 $\text{plim}(W_n) = \theta$，定义 γ 的一个估计量为 $G_n = g(W_n)$，于是

$$\text{plim}(G_n) = \gamma \tag{C.9}$$

这个结果常写为：对于一个连续函数 $g(\theta)$，我们有

$$\text{plim}\, g(W_n) = g(\text{plim}\, W_n) \tag{C.10}$$

“假定 $g(\theta)$ 是连续的”，这一技术性要求常被人们用非技术性语言描述为“它是一个能通过笔不离纸画出来的函数”。因为本书中遇到的函数都是连续的，我们就不对连续函数提供一个正式的定义了。连续函数的例子有如：$g(\theta) = a + b\theta$（a 和 b 为常数）；$g(\theta) = \theta^2$；$g(\theta) = 1/\theta$，$g(\theta) = \sqrt{\theta}$，$g(\theta) = \exp(\theta)$，等

711 等。我们将不需要再次提连续性假定了。

作为一个一致的但有偏误的估计量的重要例子，且考虑估计以 μ 为均值、σ^2 为方差的总体标准差 σ。我们曾声称样本方差 $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ 是方差 σ^2 的无偏估计。利用大数定律和一些代数，还可以证明 S^2 也是

σ^2 的一致估计。很自然，把 $\sigma=\sqrt{\sigma^2}$ 估计为 $S_n=\sqrt{S_n^2}$（其中，平方根总取它的正值）。因为平方根的期望值不是期望值的平方根（见 B.3），所以称之为**样本标准差**（sample standard deviation）的 S_n 不是一个无偏的估计量。然而，由 PLIM.1，$\text{plim}S_n=\sqrt{\text{plim}S_n^2}=\sqrt{\sigma^2}=\sigma$，因此，$S_n$ 是 σ 的一个一致估计量。

下面是概率极限的另一些有用的性质。

性质 PLIM.2

如果 $\text{plim}(T_n)=\alpha$，$\text{plim}(U_n)=\beta$，则

(i) $\text{plim}(T_n+U_n)=\alpha+\beta$

(ii) $\text{plim}(T_nU_n)=\alpha\beta$

(iii) $\text{plim}(T_n/U_n)=\alpha/\beta$，$\beta\neq 0$

这里的三个关于概率极限的等式使得我们能以种种方式把一些一致估计量组合成另一些一致估计量。例如，令$\{Y_1,\cdots,Y_n\}$为来自受过中学教育的工人总体、样本大小为 n 的一个关于年薪的随机样本，并记其总体均值为 μ_Y。再令$\{Z_1,\cdots,Z_n\}$为来自受过大学教育的工人总体的一个关于年薪的随机样本，并记其总体均值为 μ_Z。我们想估计这两类工人年薪的百分数差异，即估计 $\gamma=100\cdot(\mu_Z-\mu_Y)/\mu_Y$。（这是大学毕业生平均薪金异于中学毕业生平均薪金的百分比。）由于 $\bar{Y}_n$ 和 $\bar{Z}_n$ 分别是 μ_Y 和 μ_Z 的一致估计量，由 PLIM.1 和 PLIM.2 的（iii）推知

$$G_n\equiv 100\cdot(\bar{Z}_n-\bar{Y}_n)/\bar{Y}_n$$

是 γ 的一个一致估计量。G_n 恰恰是样本中 $\bar{Z}_n$ 和 $\bar{Y}_n$ 的百分数差异，因而它是一个很自然的估计量。G_n 虽然不是 γ 的一个无偏估计量，但仍是一个好的估计量，除非 n 很小。

渐近正态性

一致性是点估计量的一个性质。虽然它告诉我们，随着样本变大，这个估计量的分布将围绕所估参数而退化到一点，但它根本没有告诉我们这个分布在给定样本大小时的形状。为了构造区间估计量和检验假设，我们需要有一种方法来逼近我们的估计量的分布。大多数的计量经济估计量，都有一个在大样本条件下可用正态分布去良好地逼近的分布。这就是引进以下定义的动机。

渐近正态性的含义

712 令$\{Z_n:n=1,2,\cdots\}$是随机变量的这样的一个序列，使得对所有的数 z，有

$$\text{P}(Z_n\leqslant z)\rightarrow\Phi(z),\ n\rightarrow\infty \tag{C.11}$$

式中，$\Phi(z)$ 为标准正态累积分布函数。这样就说 Z_n 有一个渐近标准正态

分布。我们常把这一情形写成 $Z_n \overset{a}{\sim} \text{Normal}(0, 1)$。波浪号上方的“$a$”表示“渐近（asymptotically 或 approximately）”的意思。

性质（C.11）是说，随着样本大小 n 变大，Z_n 的累积分布函数变得越来越接近标准正态分布的 cdf。当渐近正态性成立时，对于大的 n，我们有近似式 $P(Z_n \leqslant z) \approx \Phi(z)$。就是，可以用标准正态概率去逼近 Z_n 关于的概率。

中心极限定理（central limit theorem，CLT）是概率与统计学中最强有力的结果之一。它表明任何（有有限方差的）总体的一个随机样本的均值，经过标准化后，都有一个渐近的标准正态分布。

中心极限定理

令 $\{Y_1, Y_2, \cdots, Y_n\}$ 为一个有均值 μ 和方差 σ^2 的随机样本。于是

$$Z_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \tag{C.12}$$

有一个渐近的标准正态分布。

式（C.12）中的 Z_n 变量就是 $\bar{Y}_n$ 变量的标准化形式：我们从 $\bar{Y}_n$ 减去了 $E(\bar{Y}_n) = \mu$，然后除以 $\text{sd}(\bar{Y}_n) = \sigma/\sqrt{n}$。这样，无论 Y 的总体分布是什么，Z_n 都有和标准正态分布一样的零均值和单位方差。令人惊奇的是，随着 n 变大，Z_n 的整个分布变得任意地接近于标准正态分布。

统计学和计量经济学中遇到的大多数估计量都可写为样本均值的函数，这时我们就能应用大数定律和中心极限定理。当两个一致估计量都有渐近的正态分布时，我们选择有最小渐近方差的一个估计量。

除了式（C.12）中的标准化样本均值外，还有许多取决于样本均值的其他统计量也是渐近正态的。一个重要的统计量是将方程（C.12）中的 σ 代之以它的一致估计量 S_n 而获得

$$\frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}} \tag{C.13}$$

对于大的 n 也有一个近似的标准正态分布。式（C.12）和式（C.13）的精确的（有限样本的）分布肯定不相同，但对于大的 n，其差异往往小得可以忽略不计。

718 本节中每个用 n 作为下标的变量都是为了强调渐近或大样本分析的性质。渐近分析的要义一旦清楚，再继续这样做下去，则徒然使符号变得臃肿而无益于增进理解。从现在起，我们略去下标 n，并相信你能记住估计量总是和样本大小有关的；诸如一致性和渐近正态性等性质均指样本大小在无限地增长的情形。

C.4　参数估计的一般方法

到现在为止，我们一直用样本均值来说明估计量的有限和大样本性质。我们不禁要问：有没有一般的估计方法能产生具有诸如无偏性、一致性和有效性等良好性质的估计量呢?

我们的回答是有。但详细论述各种估计方法超出了本书的范围；这里只作些非正式的讨论。透彻的讨论可参阅拉森和马克斯（Larsen and Marx, 1986，第 5 章）的论著。

矩法

给定总体中的一个参数 θ、通常都有多种方法获得 θ 的无偏和一致估计量。尝试所有不同的方法并根据 C.2 节和 C.3 节的准则去比较它们是不现实的。幸而已经证明一些方法有良好的一般性质，并且它们的合理性大多是相当直观的。

在前面各节中，我们已经看到了用**矩法**（method of moment）演算的一些例子。基本上，矩法估计是这样进行的：把参数 θ 表示成与 Y 分布中的某些期望值有某种关系，通常是与E(Y)或 E(Y^2)(虽然有时会有更独特的选择)有关系。例如，我们感兴趣的参数 θ 与总体均值 μ 有 $\theta = g(\mu)$的关系，其中 g 表示某个函数。由于样本均值 $\bar{Y}$ 是 μ 的一个无偏且一致估计量，用 $\bar{Y}$ 代替 μ 是很自然的，这样就产生了 θ 的估计量 $g(\bar{Y})$。估计量 $g(\bar{Y})$对 θ 是一致的；如果 $g(\mu)$是 μ 的线性函数，则 $g(\bar{Y})$还是无偏的。我们所做的就是用样本矩 $\bar{Y}$ 代替了相应的总体矩。“矩法”的名称便由此而来。

我们来介绍回归分析讨论中用得着的另外两个矩法估计量。回想两随机变量 X 和 Y 的协方差曾被定义为 $\sigma_{XY} = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)]$。矩法示意我们把 σ_{XY} 估计为 $n^{-1}\sum_{i=1}^{n}(X_i - X)(Y_i - \bar{Y})$，这是 σ_{XY} 的一个一致估计量，但却不是无偏的，其理由和用 n 而不用 $n-1$ 作为分母时得到的样本方差是有偏误的理由本质上是一样的。**样本协方差**（sample covariance）被定义为

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{C.14}$$

714　可以证明，这是 σ_{XY} 的一个无偏误估计量（当样本无限地增大时，用 $n-1$ 代替 n 产生不了什么差异，因此这个估计量仍是一致的）。

如在 B.4 节中讨论过的，两个变量之间的协方差常常难以解释。通常我们更关注相关性。由于总体相关是 $\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$，故矩法示意我们把 ρ_{XY} 估计为

$$R_{XY}=\frac{S_{XY}}{S_XS_Y}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]^{1/2}\left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]^{1/2}}\tag{C.15}$$

它被称之为**样本相关系数**（sample correlation coefficient）或简称样本相关。

注意我们已把样本协方差和样本标准差中的除数 $n-1$ 相消。事实上，即使我们用 n 除两者也会获得同样的最后公式。

可以证明，样本相关系数和我们认为的那样，永远位于区间 $[-1, 1]$ 上。因为 S_{XY}，S_X 和 S_Y 对相应的总体参数来说是一致的，所以 R_{XY} 是总体相关 ρ_{XY} 的一致估计量。然而，有两个理由使得 R_{XY} 是一个偏误估计量。第一，S_X 和 S_Y 分别是 σ_X 和 σ_Y 的偏误估计量；第二，R_{XY} 是一些估计量之比率，所以不是无偏的（即令 S_X 和 S_Y 无偏）。虽然在数理统计学中不存在有 ρ_{XY} 的无偏估计量已是一个经典的结果，但 R_{XY} 有偏误对我们的目的来说并不重要。

最大似然法

另一个一般性的估计方法是最大似然法，许多入门统计学课程都含有这个子目。这里只需要扼要地叙述最简单的情形。令 $\{Y_1, Y_2, \cdots, Y_n\}$ 为来自总体分布 $f(y;\theta)$ 的一个随机样本。因为假定了随机抽样，所以 $\{Y_1, Y_2, \cdots, Y_n\}$ 的联合分布无非是各个密度的乘积：$f(y_1;\theta)f(y_2;\theta)\cdots f(y_n;\theta)$。对于离散情形，这就是 $\mathrm{P}(Y_1=y_1, Y_2=y_2, \cdots, Y_n=y_n)$。现在定义**似然函数**（likelihood function）为

$$L(\theta;Y_1,\cdots,Y_n)=f(Y_1;\theta)f(Y_2;\theta)\cdots f(Y_n;\theta)\tag{C.16}$$

这是一个随机变量，因为它取决于随机样本 $\{Y_1, Y_2, \cdots, Y_n\}$ 的结果。θ 的**最大似然估计量**（maximum likelihood estimator, MLE），且叫做 W，是使似然函数最大化的 θ 值。（这就是为什么我们把 L 写成 θ 并接着是这个随机样本的一个函数的缘故）。显然，这个值取决于随机样本。

最大似然原理是说，在所有可能的 θ 值中，应选取使观测数据有最大似然性的 θ 值，直观上，这是估计 θ 的一种合理方法。

最大似然估计通常是一致性的并且有时是无偏的，然而有许多其他的估计量也是这样。MLE 之所以广为流传是因为它一般都是渐近最有效的，如果总体模型是正确地设定的话。其次，MLE 有时是**最小方差无偏估计量**
715 （minimum variance unbiased estimator）；就是说，在 θ 的所有无偏估计量中，它有最小的方差。［为证实这点，参看 Larsen 和 Marx（1986，第 5 章）。］本书只在第 3 篇的一部分高深专题中需要用到 MLE。

最小二乘法

第三类估计量并且是在本书中一直扮演着主要角色的一类估计量，是所谓**最小二乘估计量**（least squares estimator）。我们曾经看到最小二乘的一个例子：样本均值 $\bar{Y}$ 是总体均值 μ 的一个最小二乘估计量。我们已经知道 $\bar{Y}$ 是一个矩法估计量，怎么又会是一个最小二乘估计量呢？可以证明，使离差平方和

$$\sum_{i=1}^{n}(Y_i - m)^2$$

尽可能小的 m 值是 $m=\bar{Y}$。证明这点并不难，但我们略去其代数运算。

对于一些分布，包括正态和贝努里，样本均值 $\bar{Y}$ 也是总体均值 μ 的最大似然估计量。因此，最小二乘、矩法和最大似然原理常常给出同样的结果。在另一些情形中，这些估计量是类似的，但不完全相同。

C.5 区间估计与置信区间

区间估计的性质

从一个特定的样本得到的点估计，本身还不足以对检验经济理论或政策探讨提供适当信息。一个点估计也许是研究者对总体值的最好猜测，但根据其性质，并不能告诉我们估计值到底离总体参数有“多么”近。作为一个例子，假定某研究者根据工人的一个随机样本报导说，在职培训津贴使每小时工资增加6.4%。我们怎能知道一旦整个工人总体接受了培训，其效果是否接近这个数呢？由于不知道总体值，我们无法知道某一特定的估计值究竟离它多近。然而，我们能作出概率方面的陈述，并由此迎来了区间估计。

我们已经知道评价一个估计量的不确定性的一个方法：求出它的抽样标准差。连同点估计一起报导估计量的标准差，能为我们的估计值的精确度提供一些信息。然而，这个标准差依赖于未知的总体参数（指 σ^2——译者注），就算这个问题可以忽略，连同点估计一起报导标准差，也并没有直接陈述总体值很可能落在相对于估计值的什么地方。通过**置信区间**（confidence interval）的构造，克服了这一障碍。

我们用例子阐明置信区间的概念。假定总体有 Normal（μ, 1）分布，并令 $\{Y_1, \cdots, Y_n\}$ 是来自这个总体的一个随机样本。（为易于说明，我们假定了总体方差已知并等于1；然后我们再说明怎样对付方差未知的更为现实的

情形。）样本均值 $\bar{Y}$ 有一个均值为 μ、方差为 $1/n$ 的正态分布；$\bar{Y} \sim \text{Normal}(\mu, 1/n)$。现在将 $\bar{Y}$ 标准化，并因标准化的 Y 有一标准正态分布而有

$$P\left(-1.96<\frac{\bar{Y}-\mu}{1/\sqrt{n}}<1.96\right)=0.95$$

括号中的事件等同于事件 $\bar{Y}-1.96/\sqrt{n}<\mu<\bar{Y}+1.96/\sqrt{n}$，于是

$$P(\bar{Y}-1.96/\sqrt{n}<\mu<\bar{Y}+1.96/\sqrt{n})=0.95 \tag{C.17}$$

方程（C.17）是令人感兴趣的，因为它告诉了我们，随机区间 $[\bar{Y}-1.96/\sqrt{n}, \bar{Y}+1.96/\sqrt{n}]$ 包含总体均值 μ 的概率是 0.95 或 95%。这一信息使我们能构造 μ 的一个区间估计，方法是把均值 $\bar{y}$ 的样本结构代入。于是

$$[\bar{y}-1.96/\sqrt{n}, \bar{y}+1.96/\sqrt{n}] \tag{C.18}$$

给出了 μ 的一个区间估计的例子。它又被称为一个 95% 置信区间，可简记为 $\bar{y}\pm1.96/\sqrt{n}$。

一旦观测到样本数据 $\{y_1, y_2, \cdots, y_n\}$，方程（C.18）中的置信区间是容易计算的。$\bar{y}$ 是依赖于数据的惟一因子。例如，假定 $n=16$，并且 16 个数据点的均值是 7.3，则 μ 的这个 95% 置信区间是 $7.3\pm1.96/\sqrt{16}=7.3\pm0.49$，这又可写成 [6.81, 7.79]。由构造知 $\bar{y}=7.3$ 是此区间的中心。

置信区间的计算虽易，理解则较难。当我们说方程（C.18）是 μ 的一个 95% 置信区间时，我们的意思是说，随机区间

$$[\bar{Y}-1.96/\sqrt{n}, \bar{Y}+1.96/\sqrt{n}] \tag{C.19}$$

包含 μ 的概率是 0.95。换言之，在抽取随机样本之前即有 95% 的机会方程（C.19）包含 μ。方程（C.19）是**区间估计量**（interval estimator）的一个例子。它是一个随机区间，因为端点随不同的样本而变。

人们常常这样解释置信区间："μ 落在区间（C.18）上的概率是 0.95"。这种解释是错误的。一旦样本被观测到且 $\bar{y}$ 也已计算出来，则置信区间的上限、下限不过是两个数（在上述例子中就是 6.81 和 7.79）。总体参数 μ 虽然未知，也仅是个常数。因此 μ 或者落入或者不落入区间（C.18）（我们
717 永远不会确切知道哪一情形是对的）。一旦利用现有的数据把置信区间计算出来，就不再有概率的问题。概率意义出自如下事实：对所有随机样本中的 95% 来说，这样构造的置信区间将包含 μ。

为了强调置信区间的意义，在表 C.2 列出取自 Normal（2.1）分布的大小为 $n=10$ 的 20 个随机样本（或重复试验）的计算结果，对 20 个样本中的每一个求出 $\bar{y}$，并把方程（C.18）计算为 $\bar{y}\pm1.96/\sqrt{10}=\bar{y}\pm0.62$（四舍五入到两位小数）。你可以看到，每个随机样本的区间在变。20 个区间中的 19 个包含了总体 μ 值。只在第 19 次重复试验中 μ 不落在置信区间里。换句话说，有 95% 的样本给出了含有 μ 的置信区间。对于只有 20 次重复的情形，这个结果不是必然的，但在这个特定的仿真实验里却得到了这一结果。

表 C.2　取自 Normal $(\mu,1)$，其中 $\mu=2$ 的分布的仿真置信区间

重复	y	95%区间	包含 μ 吗？
1	1.98	(1.36,2.60)	是
2	1.43	(0.81,2.05)	是
3	1.65	(1.03,2.27)	是
4	1.88	(1.26,2.50)	是
5	2.34	(1.72,2.96)	是
6	2.58	(1.96,3.20)	是
7	1.58	(0.96,2.20)	是
8	2.23	(1.61,2.85)	是
9	1.96	(1.34,2.58)	是
10	2.11	(1.49,2.73)	是
11	2.15	(1.53,2.77)	是
12	1.93	(1.31,2.55)	是
13	2.02	(1.40,2.64)	是
14	2.10	(1.48,2.72)	是
15	2.18	(1.56,2.80)	是
16	2.10	(1.48,2.72)	是
17	1.94	(1.32,2.56)	是
18	2.21	(1.59,2.83)	是
19	1.16	(0.54,1.78)	否
20	1.75	(1.13,2.37)	是

正态分布总体均值的置信区间

从方程（C.18）导出置信区间有助于说明如何去构造并解释置信区间。其实，为了求一个正态总体均值，方程（C.18）由于假定了已知方差为 1，就不是很有用的。但容易把方程（C.18）推广到标准差 σ 为任意已知值的情形：95%置信区间是

$$[\bar{y}-1.96\delta/\sqrt{n},\bar{y}+1.96\delta/\sqrt{n}] \tag{C.20}$$

因此，当 σ 已知时，μ 的置信区间便已构造出来。为了考虑 σ 求值的情形，我们必须用一个估计值。令

$$s=\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]^{1/2} \tag{C.21}$$

表示样本标准差。于是，用σ的估计值S代替式（C.20）中的σ，我们就能求出一个完全依赖于观测数据的置信区间。麻烦的是，这样一来，由于s依赖于特定样本，就不能保持95%的置信水平。换句话说，因为常数σ已被随机变量s代替，所以随机区间$[\bar{Y}\pm1.96\,(S/\sqrt{n})]$包含$\mu$的概率不再是0.95。

我们该怎样做下去呢？我们必须依靠t分布，而不是使用标准正常分布。t分布出自

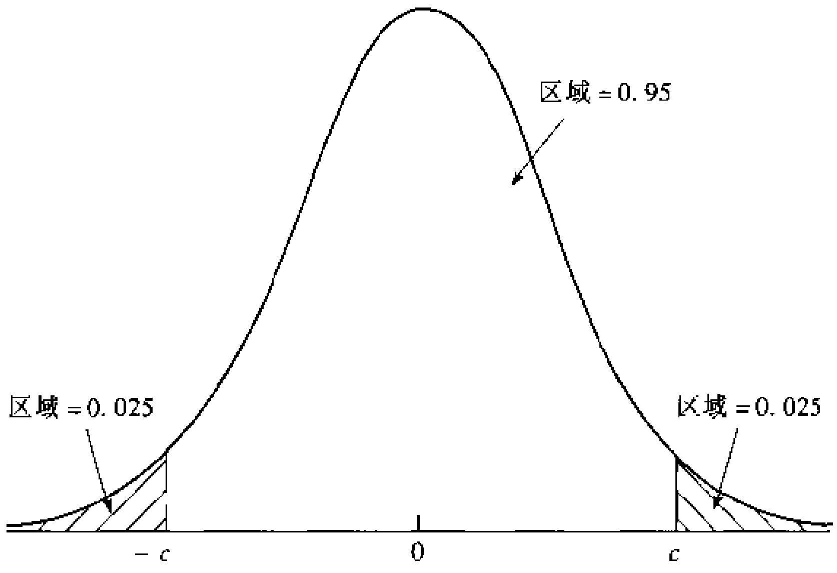
$$\frac{\bar{Y}-\mu}{S/\sqrt{n}}\sim t_{n-1} \tag{C.22}$$

719 式中，$\bar{Y}$为样本均值；S为随机样本$\{Y_1,\cdots,Y_n\}$的样本标准差。我们将不去证明式（C.22）；在许多文献中都可找到它的细致的证明［例如，Larsen和Marx（1988，第7章）］。

为了构造一个95%置信区间，令c表示t_{n-1}分布中的第97.5分位数。换言之，c是这样一个值，使得t_{n-1}中的面积有95%落在$-c$与c之间：$P(-c<t_{n-1}<c)=0.95$。（c值依赖于自由度$n-1$，但我们没有标明。）图C.4说明了c的选择。一旦适当选定了c，随机区间$[\bar{Y}-c\cdot S/\sqrt{n},\ \bar{Y}+c\cdot S\sqrt{n}]$包含了$\mu$的概率就是0.95。对于一个特定的样本，这个95%置信区间将计算为

$$[\bar{y}-c\cdot s/\sqrt{n},\bar{y}+c\cdot s\sqrt{n}] \tag{C.23}$$

对不同自由度的c值可从附录G的表G.2查得。例如，当$n=20$，因而自由度是$n-1=19$，$c=2.093$。于是，95%置信区间是$[\bar{y}\pm2.093\,(s/\sqrt{20})]$，其中$\bar{y}$和$S$都是从样本算出的值。即使$s=\sigma$（很少可能），式



图C.4　t分布中的第97.5百分位数c

（C.23）中的置信区间由于 $c>1.96$ 也要比式（C.20）中的宽些。自由度越小，式（C.23）越宽。

720 更一般地，令 c_α 表示 t_{n-1} 分布中的第 100（$1-\alpha$）百分位数，则可把 100（$1-\alpha$）%置信区间取为

$$[\bar{y}-c_{\alpha/2}s/\sqrt{n}, \bar{y}+c_{\alpha/2}s/\sqrt{n}] \tag{C.24}$$

要得到 $c_{\alpha/2}$，首先要选择 α 并知道自由度 $n-1$，然后才能使用表 G.2。大多数情形用的都是 95%置信区间。

有一个简单的方法去记住怎样构造一个正态分布的均值的置信区间。回忆 $\mathrm{sd}(\bar{Y})=\sigma/\sqrt{n}$，而 $s/\sqrt{n}$ 是 $\mathrm{sd}(\bar{Y})$ 的点估计(值)，于是有时把相应的随机变量 $S/\sqrt{n}$ 叫做 $\bar{Y}$ 的**标准误**（standard error）。因为出现在公式中的是点估计 $s/\sqrt{n}$，我们就把 $\bar{y}$ 的标准误定义为 $\mathrm{se}(\bar{y})=s/\sqrt{n}$。于是式（C.24）可简写为

$$[\bar{y}\pm c_{\alpha/2}\cdot \mathrm{se}(\bar{y})] \tag{C.25}$$

这一方程表明了为什么一个估计值的标准误含义在计量经济学中扮演着重要的角色。

例 C.2　在职培训津贴对工人生产力的影响

霍尔泽、布罗克、奇塔姆和诺特（Holzer, Block, Cheatham and Knott, 1993）通过对 1988 年领取在职培训津贴的密歇根州制造业厂商的一个样本收集的废弃率信息，研究了在职培训津贴对工人生产力的影响。表 C.3 列出了 20 家厂商的废弃率。废弃率被度量为每生产 100 件不能使用而必须报废的件数。这些厂商都在 1988 年领取了在职培训津贴；1987 年还没有津贴奖励。我们的兴趣在于，对 1987 年到 1988 年废弃率的改变构造一个置信区间，以估计如果全部制造厂商作为一个总体都领到了津贴，总体废弃率如何。

表 C.3　　20 家密歇根州制造厂商的废弃率

厂商	1987 年	1988 年	改变
1	10	3	-7
2	1	1	0
3	6	5	-1
4	0.45	0.5	0.05
5	1.25	1.54	0.29
6	1.3	1.5	0.2
7	1.06	0.8	-0.26
8	3	2	-1
9	8.18	0.67	-7.51
10	1.67	1.17	-0.5
11	0.98	0.51	-0.47

721

12	1	0.5	0.5
13	0.45	0.61	0.16
14	5.03	6.7	1.67
15	8	4	4
16	9	7	2
17	18	19	1
18	0.28	0.2	0.08
19	7	5	−2
20	3.97	3.83	−0.14
平均	4.38	3.23	1.15

假定废弃率的改变是正态分布的。因 $n=20$，故废弃率改变的均值 μ 的一个 95% 置信区间是 $[y \pm 2.093 \cdot \mathrm{se}(\bar{y})]$，其中 $\mathrm{se}(\bar{y})=s/\sqrt{n}$。数值 2.093 是 t_{19} 分布中的第 97.5 百分位数。对这个特定样本数据，$y=-1.15$ 和 $\mathrm{se}(\bar{y})=0.54$（这两个数都四舍五入到两位小数），因此得 95% 置信区间为 $[-2.28, -0.02]$。

由于此区间不包括零值，所以我们的结论是，有 95% 的置信度，总体废弃率改变的均值 μ 不是零。

722 在这个问题上，例 C.2 主要用于说明方法。从计量经济分析来说，这个例子隐含有一些严重的缺陷。最重要的，它假定了废弃率的任何系统下降都是由于在职培训津贴的作用。但是在这一年里，许多事情的发生都会与工人生产力变化有关。从这一分析看，我们无法知道平均废弃率的下降是否可归功于在职培训津贴，抑或，至少部分地，归因于一些外部力量。

95%置信区间的一个简单的经验法则

对任何样本大小和任何置信水平都可计算式（C.25）中的置信区间。如我们在 B.4 节所看到的，t 分布随着自由度的增大而接近于标准正态分布。特别地，对于 $\alpha=0.05$，随着 $n \to \infty$，而有 $c_{\alpha/2} \to 1.96$，尽管对于每个 n，$c_{\alpha/2}$ 总比 1.96 大。作为求一个近似的 95% 置信区间的经验法则是取它为

$$[\bar{y} \pm 2 \cdot \mathrm{se}(\bar{y})] \qquad (C.26)$$

换言之，我们求 $\bar{y}$ 和它的标准误，再计算 $\bar{y}$ 加减 2 倍于它的标准误，便得到这个置信区间。对于很大的 n，它稍宽了一些，而对于很小的 n，它又窄了一些。如我们从例 C.2 能看到的，即使 n 小到 20，式（C.20）仍可用做正态分布均值的一个 95% 置信区间。就是说，为获得一个相当接近 95% 的置信区间，大可不必参照 t 分布表。

非正态总体的渐近置信区间

在一些应用中，总体明显不是正态的。一个重要的例子是随机变量只取 0 和 1 两个值的贝努里分布。其他的非正态总体则没有标准的分布。但如果样本大到使得中心极限定理能给出样本均值 $\bar{Y}$ 的分布一个良好近似的话，怎样的非正态都无关重要了。对于大的 n，一个近似的 95%置信区间是

$$[\bar{y} \pm 1.96 \cdot \mathrm{se}(y)] \tag{C.27}$$

式中的数值 1.96 是标准正态分布的第 97.5 百分位数。机械地看，计算一个近似置信区间无异于正态的情形。一个轻微的差异在于标准误的倍数来自标准正态分布而不是 t 分布，因为我们是在引用渐近理论。由于 t 分布随 df 增加而接近标准正态，方程（C.25）作为近似 95%区间乃是完全合法的，一些人认为方程（C.25）比方程（C.27）更可取还因为前者对于正态总体来说是一个精确的结果。

723

例 C.3 雇用中的种族歧视

1988 年华盛顿特区的城市管理部门做了一项调查，研究在雇用员工中种族歧视的程度。在几种工作上面试了 5 对申请人。每一对申请人中都有一个黑人和一个白人。他们的简历都表明他们有差不多的工作经验、教育水平以及决定工作资历方面的类似的其他因素，用意是要使除种族外两个人尽可能相似。

参加面试的一对申请者都申请相同的工作，研究者记下哪一申请者被录用。这是所谓配对分析（matched pairs analysis）的一个例子。在这里，每对被认为除了一个重要特征外各个方面都相似的人（厂商、城市等等）的数据构成一次试验。

令 θ_B 代表黑人被录用的概率，θ_W 代表白人被录用的概率。我们关注的主要是差值 $\theta_B - \theta_W$。令 B_i 代表如果雇主 i 录用黑人就等于 1，否则等于 0 的贝努里变量。类似地，如果雇主 i 录用白人则 $W_i = 1$，否则 $W_i = 0$。把 5 对申请人对所有雇主统计起来共有 $n = 241$ 次试验（241 对雇员的面试）。θ_B 和 θ_W 的无偏估计量是 $\bar{B}$ 和 $\bar{W}$，分别代表面试中黑人和白人被录用的分数。

为把上述各点纳入计算总体均值的置信区间这个框架里，我们定义一个新的变量 $Y_i = B_i - W_i$。现在，Y 可以取三个值：如果黑人不被录用而白人被录用，则 Y 取值 -1；如果两人都被录用或都不被录用，则 Y 取值 0；如果黑人被录用而白人不被录用，则 Y 取值 1。于是，$\mu \equiv \mathrm{E}(\bar{Y}_i) = \mathrm{E}(B_i) - \mathrm{E}(W_i) = \theta_B - \theta_W$。

Y_i 的分布肯定不是正态的——它既是离散的，而又仅取三个值。然而，利用大样本方法可以得到 $\theta_B - \theta_W$ 的一个近似置信区间。

使用这 241 个观测数据点，$\bar{b} = 0.224$ 而 $\bar{w} = 0.357$，因此 $\bar{y} = 0.224 -$

$0.357 = -0.133$。就是说，有 22.4% 的黑人申请者被录用，而白人申请者有 35.7% 被录用。这是歧视黑人的真凭实据。然而，通过计算 μ 的一个置信区间，我们可以知道更多的东西。为了计算一个近似的 95% 的置信区间，我们需要有样本标准差。计算的结果是 $s = 0.482$［利用方程(C.21)］。利用方程(C.27)我们得到 $\mu = \theta_B - \theta_W$ 的一个 95% 置信区间为 $-0.133 \pm 1.96(0.482/\sqrt{241}) = -0.133 \pm 0.031 = [-0.164, -0.102]$。近似的 99% 置信区间是 $-0.133 \pm 2.58(0.482/\sqrt{241}) = [-0.213, -0.053]$。这自然比 95% 置信区间包含着一个更宽的数值范围。但即使是 99% 置信区间，它也不包含零值。因此，我们非常有信心地说，总体差值 $\theta_B - \theta_W$ 不是零。

在我们离开置信区间之前，还需要作最后一点评注。因为 $\bar{y}$ 的标准误 $\mathrm{se}(\bar{y}) = s/\sqrt{n}$ 随着样本大小的增加而趋于零，我们看到——其他条件不变下——较大的样本意味着较小的置信区间。因此，大样本的一个重大好处在于它导致了较小的置信区间。

C.6 假设检验

724 至今，我们复习了怎样去评价点估计量，并且看到了——对于总体均值的情形——怎样去构造并解释置信区间。但有时我们感兴趣的问题需要有一个肯定是或不是的答案。这里给出几个例子：(1) 一个在职培训计划有效地增加了平均的工人生产力吗？(参看例 C.2)；(2) 黑人在员工雇用中受到歧视吗？(参看例 C.3)；(3) 更严厉的州酒后驾驶法减少了因酒后驾驶而被逮捕的次数吗？设计利用样本数据来回答这种问题的方法就名为假设检验。

假设检验的基本知识

为了说明有关假设检验的问题，考虑一个选举问题。假定在一次选举中有两个候选人 A 和 B。据报导，候选人 A 已得到 42% 的选票，而候选人 B 得到 58% 的选票。姑且把这个百分比看成选民总体的真正百分比。

候选人 A 深信更多的民众会投他的票，因此想调查选举是否有作弊情况。知道一点统计学的候选人 A，雇用了一个咨询机构去随机地抽取 100 名选举人的一个样本，记录了每名选举人是否投他的票。假定在所收集的样本中，有 53 人投了赞成候选人 A 的票。这一样本估计值 53% 明显超过所报告的总体值 42%。候选人 A 是否应作出结论说选举曾是一个骗局呢？

虽然表面看来候选人 A 的得票率被低估了，但我们还不能肯定。即使只有总体中 42% 的人赞成了候选人 A 是真的，在一个 100 人的样本中仍有

可能观测到53人投了A的赞成票。问题是这个和官方报导的百分比42%相违的样本证据到底有多强。

处理这个问题的一个方法，是设立一个**假设检验**（hypothesis test）。令θ代表赞成候选人A的总体真实比例，令所报告的结果为真实的假设，可陈述为

$$H_0: \theta = 0.42 \tag{C.28}$$

这是**虚拟假设**（null hypothesis）的一个例子。我们总是把虚拟假设记为H_0。在假设检验中，虚拟假设扮演的角色，类似于许多司法制度中的案件被控告人：好比被告人在没有被证明有罪之前就是清白无辜的，虚拟假设在有强有力的数据表明它有问题之前就是真实的。在现在这个例子中，候选人A必须拿出相当强的反对假设（C.28）的证据，才能赢得一次重新计票。

在选举例子中，**对立假设**（alternative hypothesis）是选举中赞成候选人A的真实比例大于0.42：

$$H_1: \theta > 0.42 \tag{C.29}$$

为了能作出H_0错误而H_1正确的结论，我们必须提出反对H_0的“合理的毋庸置疑”的证据。在我们感到反对H_0的证据是强有力的之前，我们需要在100张选票中有多少是赞成A的呢？大多数人都会同意，在大小为100的样本中只观测到43张赞成票是不足以推翻原始选举结果的；这样的结果完全 725 属于预料的抽样波动范围。另一方面，我们也不需要观测到有了100张赞成A的票再对H_0提出怀疑。100人中有53人赞成是否足以拒绝H_0就不很清楚了。答案有赖于怎样量化“合理的毋庸置疑”。

在假设检验中我们会犯两种错误：首先，我们可能拒绝一个其实是真的虚拟假设。这叫做第Ⅰ类错误。在选举例子中，当赞成候选人A的人的比例事实上是0.42时而我们拒绝了H_0，就出现第Ⅰ类错误。第二种错误是指H_0实际上是错误的，但我们并没有拒绝它。这叫做第Ⅱ类错误。在选举例子中，如果$\theta>0.42$而我们没有拒绝H_0，就出现第Ⅱ类错误。

在我们作出是否拒绝虚拟假设的决定之后，我们不是做了正确的决定就是犯了一种错误。我们永远不会确切知道是否犯了错误。然而，我们能计算出或者是犯了第Ⅰ类错误或者是犯了第Ⅱ类错误的概率。人们把假设检验的规则构成使得犯第Ⅰ类错误的概率相当地小。通常我们把一个检验的**显著（性）水平**（significance level）[或简称检验的水平（或大小）——检验的水平（level）又称检验的大小（size）——译者注]定义为第Ⅰ类错误的概率，并典型地记为α，用符号表示就是

$$\alpha = P(\text{拒绝 } H_0 \mid H_0) \tag{C.30}$$

等式的右端读作“当H_0为真实时拒绝H_0的概率”。

经典的假设检验要求我们，在做一个检验时，一开始就定下一个α值。当设定了α值，就基本上量化了我们对第Ⅰ类错误的容忍度。通常α的值有0.10，0.05和0.01。如果$\alpha=0.05$，则研究者为了侦察出对H_0的偏离，

愿意以 5% 的机会错误地拒绝 H_0 而行事。

一旦选定显著水平，我们就想把第Ⅱ类错误的概率减到最小。或者说，我们想对所有有意义的对立情况使一个**检验的功效**（power of a test）最大。一个检验的功效不外是 1 减去第Ⅱ类错误的概率。数学上表示为

$$\pi(\theta)=\mathrm{P}(\text{拒绝 } \mathrm{H}_0 \mid \theta)=1-\mathrm{P}(\text{第Ⅱ类错误} \mid \theta)$$

式中，θ 为参数的真实值。自然，我们希望每当虚拟假设是错误的时候，这个功效能等于 1。但为了保持小的显著水平，这是办不到的。作为一种替代方法，我们把检验选择为在给定显著水平下功效最大化。

检验关于正态总体均值的假设

为了相对于一个对立假设而去检验一个虚拟假设，我们需要挑选一个检验统计量（或简称统计量）和一个临界值。统计量和临界值是根据方便和给定检验的显著水平下的功效最大化愿望来选择的。在本小节中我们复习怎样去检验关于正态总体的均值的那些假设。

726 一个**检验统计量**（test statistic），记为 T，是指随机样本的某个函数。当我们对某个特定结果计算统计量时，就得到这个统计量的一个结果，我们将它记为 t。

给定一个统计量，我们便能定义一个拒绝规则，来决定什么时候我们舍弃 H_0 而选取 H_1。在本教材中，所有的拒绝规则都是拿一个检验统计量的值 t 来同一个**临界值**（critical value）c 做比较作为依据的。所有导致拒绝虚拟假设的 t 值的全体名之为**拒绝域**（rejection region）。为了定出临界值，我们要首先决定检验的显著水平。于是，给定 α，相应于 α 的临界值就由 H_0 被假定为正确时的 T 的分布来决定。我们将此临界值写为 c，而略去其与 α 有关的事实。

检验来自一个 Normal（μ，σ^2）总体的关于均值 μ 的假设是容易的。把虚拟假设表述为

$$\mathrm{H}_0:\mu=\mu_0 \tag{C.31}$$

式中，μ_0 为我们设定的值。在多数应用中 $\mu_0=0$，但不一定是零的一般情形也没有更多的困难。

怎样选择拒绝规则仍与对立假设的性质有关。人们感兴趣的三种对立假设是

$$\mathrm{H}_1:\mu>\mu_0 \tag{C.32}$$

$$\mathrm{H}_1:\mu<\mu_0 \tag{C.33}$$

和

$$\mathrm{H}_1:\mu\neq\mu_0 \tag{C.34}$$

方程（C.32）给出一个**单侧对立**（one-sided alternative），方程（C.33）也这样。当对立假设是方程（C.32）时，虚拟假设实际上是 H_1：$\mu\leqslant\mu_0$，因为只有当 $\mu>\mu_0$ 时我们才会拒绝 H_0。如果我们对 μ 值感兴趣只因 μ 至少不比 μ_0 小，那么单侧对立（C.32）是适当的。方程（C.34）是一个**双侧对立**（two-sided alternative），它是可接受的，如果我们对任何偏离虚拟假设的情况都感兴趣的话。

先考虑方程（C.32）中的对立情况。直觉上，当样本均值 $\bar{y}$ “足够”地大于 μ_0 时，我们便应拒绝 H_0 而接受 H_1。但我们应怎样确定 $\bar{y}$ 已大到足以在选定的显著水平上拒绝 H_0 呢？这要求我们知道当虚拟假设正确时而把它拒绝的概率。我们不直接同 $\bar{y}$ 打交道，而代之以它的标准化形式，并用样本标准差 s 代替其中的 σ。

$$t=\sqrt{n}(\bar{y}-\mu_0)/s=(\bar{y}-\mu_0)/\mathrm{se}(\bar{y}) \tag{C.35}$$

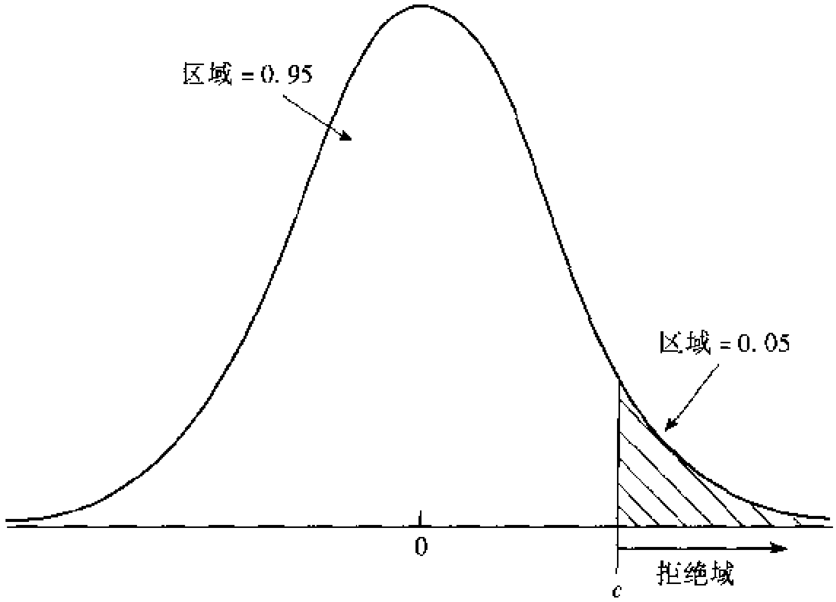
式中，$\mathrm{se}(\bar{y})=s/\sqrt{n}$ 为 $\bar{y}$ 的标准误。给定样本数据，t 是容易求出的。我们同 t 打交道的理由是，在虚拟假设下，随机变量

$$T=\sqrt{n}(\bar{Y}-\mu_0)/S$$

727 有一个 t_{n-1} 分布。现假定我们同意使用5%显著水平，于是临界值 c 的选择将使得 $\mathrm{P}(T>c\mid\mathrm{H}_0)=0.05$，就是说，第Ⅰ类错误的概率是5%。一旦我们求出 c，拒绝规则就是

$$t>c \tag{C.36}$$

式中，c 为 t_{n-1} 分布中的第100$(1-\alpha)$百分位数；作为一个百分数的显著水平将是 $100\cdot\alpha\%$。这是一个**单尾检验**（one-tailed test）的例子，因为其拒绝域位于 t 分布的一尾；图C.5说明了这点。不同的显著水平会将导致不同的临界值。



图C.5　相对于单侧对立假设的5%显著水平的拒绝域

方程（C.35）中的统计量常称为检验 $H_0: \mu=\mu_0$ 的 t 统计量。t 统计量是对于 $\bar{y}$ 的标准误 se（$\bar{y}$）的、y 距离 μ_0 的一个度量。

例 C.4　企业特区对工商业投资的影响

在某特定州被授予企业特区的城市总体中［关于印第安纳州，可参看 Papke（1994）］，令 Y 代表一个城市在成为企业特区之前的一年到在此之后的一年里投资额的百分比变化。假定 Y 有一 Normal（μ，σ^2）分布。企业特区对工商业投资没有影响的虚拟假设是 $H_0: \mu=0$，而有正的积极影响的对立假设是 $H_1: \mu>0$（假定作为特区不会出现负的消极影响）。假定我们想在 5% 水平上检验 H_0。这时，检验统计量是
728

$$t=\frac{\bar{y}}{s/\sqrt{n}}=\frac{\bar{y}}{\mathrm{se}(\bar{y})} \tag{C.37}$$

假定我们有 36 个城市被授予企业特区的一个样本，查表知临界值 $c=1.69$（见表 G.2）。因而，如果 $t>1.69$，我们就拒绝 H_0 而选取 H_1。假定样本给出 $\bar{y}=8.2$ 和 $s=23.9$。由此得 $t\approx2.06$，从而在 5% 水平上拒绝 H_0。就是说，我们的结论是在 5% 水平上，企业特区对平均投资有效果。1% 的临界值是 2.44，所以还不能在 1% 水平上拒绝 H_0。这里存在和例 C.2 同样的缺陷：我们没有在这段时间里控制好可能影响城市投资的其他因素，所以还不能说我们找到了因果性的效应。

对单侧对立（C.33）的拒绝规则是类似的。相对于方程（C.33）的有 100·α% 显著水平的一个检验是，每当

$$t<-c \tag{C.38}$$

时拒绝 H_0［而选取（C.33）］。换句话说，我们在寻找那些离开零足够远的 t 统计量的负值——这意味着 $y<\mu_0$——以拒绝 H_0。

至于双侧对立，我们必须小心选择临界值，以使检验的显著水平仍是 α。如果 H_1 由 $H_1: \mu\neq\mu$ 给出，则我们在绝对值意义上看到 y 远离 μ_0 时拒绝 H_0：一个比 μ 大得多或小得多的 $\bar{y}$，都为拒绝 H_0 而选取 H_1 提供见证。下面的拒绝规则给出一个 100·α% 显著水平的检验：

$$|t|>c \tag{C.39}$$

式中，$|t|$ 是方程（C.35）中 t 统计量的绝对值。这是一个**双尾检验**（two-tailed test）。这时我们必须小心选择临界值：c 是 t_{n-1} 分布中的第 100（$1-\alpha/2$）百分位数。例如 $\alpha=0.05$ 时，这个临界值是 t_{n-1} 分布中的第 97.5 百分位数，它保证当 H_0 正确时它只有 5% 的机会被拒绝（参看图 C.6）。例如，取 $n=22$，则临界值为 $c=2.08$，这是 t_{21} 分布中的第 97.5 百分位数（见表 G.2）。为了在 5% 水平上拒绝 H_0 而选取 H_1，t 统计量的绝对值必须超过 2.08。

熟悉假设检验的适当用语是重要的。有时人们把一个适当的词句“我们

未能在 5% 显著水平上拒绝 H_0 而选取 H_1" 换成"我们在 5% 水平上接受 H_0"。可是后者的措词是欠正确的。基于同样的数据集，常常有许多的假设是不能拒绝的。拿前面的选举例子来说，如果说 $H_0: \theta = 0.42$ 和 $H_0: \theta = 0.43$ 都"被接受"，那就出现逻辑上的矛盾，因为只能其中之一是正确的。但是两者都不被拒绝则是完全可能的。由于这个缘故，我们常说"未能拒绝 H_0"，而不说"接受 H_0"。

729

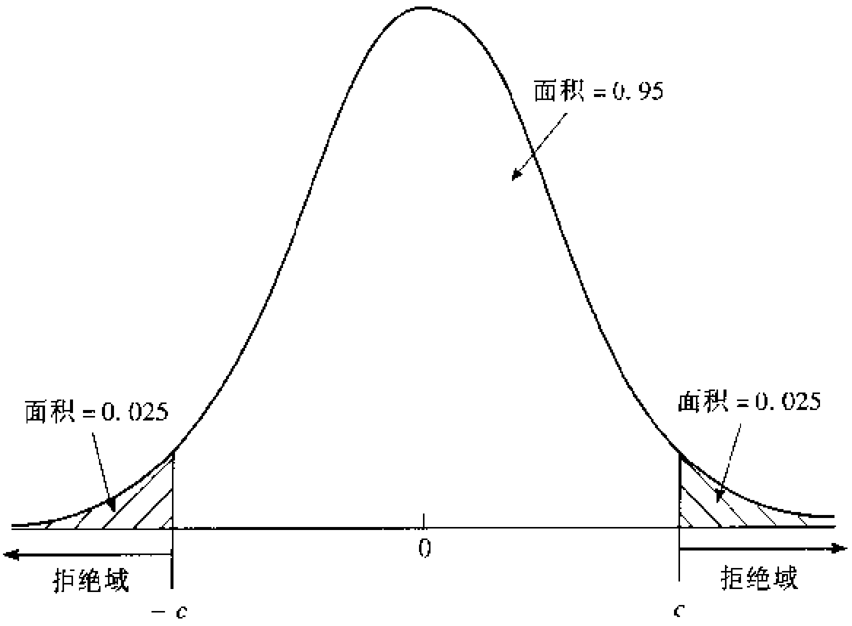


图 C.6　相对于双侧对立的 5% 显著水平检验的拒绝域

非正态总体的渐近检验

如果样本大小大到足以引用中心极限定理时（参看 C.3 节），则对总体均值进行假设检验的操作程序无论总体分布是否正态都是一样的。其理论依据来自于如下事实：在虚拟假设下，有

$$T = \sqrt{n}(\bar{Y} - \mu_0)/S \overset{a}{\sim} \text{Nomal}(0,1)$$

因此，对于大的 n，我们可以将式（C.35）中的 t 统计量和来自标准正态分布的临界值相比较。因为随着 n 变大，t_{n-1} 分布收敛于标准正态分布，所以对于极其大的 n，t 临界值和标准正态临界值是非常接近的。由于渐进理论
730 是以 n 无限增大为依据的，它不能告诉我们到底是标准正态临界值好或是 t 临界值好。对于中等大小的 n 值，比方说 30 − 60，传统习惯使用 t 分布，因为我们知道对于正态总体这样做是正确的。对于 $n > 120$，如何在分布和标准正态分布之间进行选择就不置可否了，因为这时两者的临界值实际上是相同的。

因为对非正态总体来说，无论用标准正态或 t 分布来选取临界值都仅仅

近似地生效，我们所选的临界值也仅有近似意义，所以对非正态总体来说我们的显著水平实际上是渐近的显著水平。就是说，如果我们选择一个 5% 显著水平，而我们的总体是非正态的，那么真实的显著水平将会大于或小于 5%（而且我们不知道属于哪一种情形）。当样本大小足够大时，实际显著水平将会非常接近 5%。从实践的角度说，这个差别并不重要，因此，我们现在就把形容词“渐近”省略掉。

例 C.5　员工雇用中的种族歧视

在城市管理部门对雇用中的歧视进行研究的例子（见例 C.3）中，我们主要感兴趣的是检验 H_0：$\mu=0$，其对立假设是 H_1：$\mu<0$，其中 $\mu=\theta_B-\theta_W$ 是黑人和白人被雇用的概率之差。记得 μ 是变量 $Y=B-W$ 的总体均值，其中 B 和 W 都是二值标示变量。通过对 $n=241$ 对申请者的比较，我们曾得到 $y=-0.133$ 和 $\mathrm{se}(y)=0.482/\sqrt{241}\approx0.031$。用以检验 H_0：$\mu=0$ 的 t 统计量的值是 $t=-0.133/0.031\approx-4.29$。你会记得，由附录 B，标准正态分布和有了 240 个自由度的 t 分布实际上是没有什么区别的。数值位于分布左尾的如此远方，以致我们可在任何合理的显著水平上拒绝 H_0。事实上，0.005（1%的一半）临界值（对单侧检验而言）约是 -2.58。一个等于 -4.29 的 t 值是有利于 H_1 而不利于 H_0 的强有力证据。因此，我们的结论是雇用中存在着种族歧视。

p 值的计算和使用

传统上要求事先选定一个显著性水平是因为不同的研究者用同样的数据和用同样的方法去检验同样的假设，会最终作出不同的结论。先报告我们所进行的检验的显著水平在一定程度上解决了这个问题，可是又未能完全解决。

为提供更多的信息，我们可以问问以下的问题。在最大多大的显著水平上我们能够完成一个检验并且仍未能拒绝虚拟假设？这个平均值就是我们所知的检验的 p 值。有时则称概率值。和事先选下一个显著水平并算出临界值相比，计算 p 值是较为麻烦的。但随着计算越来越快和越来越便宜，p 值现在也较易获得。

作为一个说明，考虑在一个 Normal（μ，σ^2）总体中检验 H_0：$\mu=0$ 的问题。我们利用统计量 $T=\sqrt{n}\cdot\bar{Y}/S$ 并假定 n 大到足以把 T 看做在 H_0 下有一标准正态分布。假使我们的样本给出 T 的观测值是 $t=1.52$（注意我们省略掉选择显著水平的步骤）。既然已经知道了 t 值，就能找出我们尚不能拒绝 H_0 的最大显著水平。这就是以 t 为临界值的相应显著水平。因为我们的
731 检验统计量 T 有一个在 H_0 下（意为在正确假设 H_0 下）的标准正态分布，

所以我们有

$$p\text{ 值} = P\ (T > 1.52 \mid H_0) = 1 - \Phi\ (1.52) = 0.065 \qquad (\text{C}.40)$$

式中，$\Phi(\cdot)$ 为标准正态 cdf。换言之，本例中的 p 值就是在一个正态分布中检验统计量的观测值 1.52 以右的面积。参看图 C.7 中的说明。

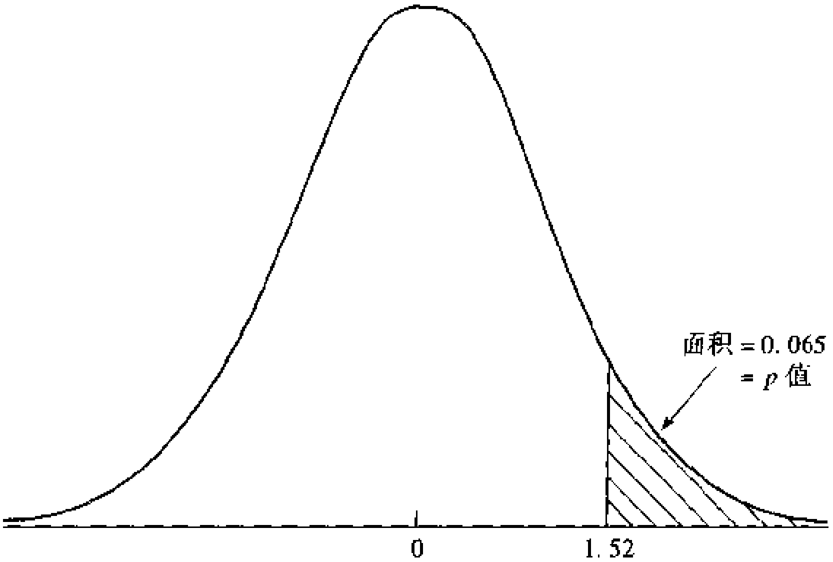


图 C.7　对于单侧对立 $\mu > \mu_0$，当 $t = 1.52$ 时的 p 值

由于 p 值 = 0.065，我们能够在完成检验后而不至于拒绝 H_0 的最大显著水平是 6.5%。如果我们在低于 6.5% 的水平（比方说 5%）上完成检验，仍不至于拒绝 H_0。但如果我们在大于 6.5% 的水平（比方说 10%）上完成检验，就会拒绝 H_0。有了 p 值，我们可以在任何水平上进行检验。

还可对本例中的 p 值作另一种有用的解释：它是在虚拟假设为正确时我们观测的 T 值大到 1.52 的概率。如果虚拟假设确实是对的，我们只有 6.5% 的机会观测到一个大到 1.52 的 T 值。这个机会是否小到足以拒绝 H_0，乃与我们对第Ⅰ类错误的容忍度有关。如我们将要看到的，对于所有的其他情形，p 值都有类似的解释。

732 一般而言，小的 p 值都是对 H_0 有疑问的证据，因为它表明了在 H_0 是真的情况下，出现这样的数据结果的概率是小的。在前面的例子中，如果 t 是较大的一个 t，比方说 $t = 2.85$，那么 p 值将是 $1 - \Phi\ (2.85) \approx 0.002$。就是说，如果虚拟假设真实的话，我们观测到一个大到 2.85 的 T 值的概率是 0.002。怎样解释这点呢？如果不是我们得到了一个很不平常的样本，那么虚拟假设就是错误的。除非我们对第Ⅰ类错误几乎不能容忍，都会拒绝虚拟假设。另一方面，较大的 p 值则是对 H_0 怀疑的弱证据。假使在上例中我们得到的是 $t = 0.47$，那么 p 值 $= 1 - \Phi\ (0.47) = 0.32$。即使当 H_0 是真的时，也有 0.32 的概率观测到一个比 0.47 大的 T 值。这个概率如此之大，要怀疑 H_0 是证据不足的，除非我们对第Ⅰ类错误有一个非常高的容忍度。

为了能用 t 分布去检验关于总体均值的假设，我们需要一个详细的数值

表以计算 p 值。表 G.2 仅允许我们算出 p 值的上、下界。幸而，现在有许多统计学和计量经济学软件包能例行地计算 p 值，并且还提供用以计算 p 值的 t 和其他分布的 cdf 计算。

例 C.6　在职培训津贴对工人生产力的影响

再考虑例 C.2 中霍尔格（Holger et al.，1993）的数据。从政策角度看，有两个有兴趣的问题。首先，什么是废弃率平均变化 μ 的最好估计值？我们已对表 C.3 中所得到的 20 个厂商的样本作了这个估计：废弃率变化的样本均值是 -1.15。相对于原来 1987 年的平均废弃率，这表示废弃率下降了约 26.3%（$-1.15/4.38 \approx -0.263$），这一效果并非微不足道。

我们也许还想知道，样本是否对制造业厂商的总体提供了强有力的证据，说明了如果它们都得到津贴也将产生积极效果。于是，虚拟假设是 H_0：$\mu=0$，而对立假设是 H_1：$\mu<0$，其中 μ 是废弃率的平均变化。我们要相对于 H_1 检验 H_0。在 H_0 下，在职培训津贴对平均废弃率无影响。而 H_1 时表示有积极影响。我们不关心另一种可能性 $\mu>0$；所以虚拟假设实际上是 H_0：$\mu \geqslant 0$。

由于 $\bar{y} = -1.15$ 和 $\mathrm{se}(\bar{y}) = 0.54$，故 $t = -1.15/0.54 = -2.13$。这低于 5% 临界值 -1.73（来自一个 t_{19} 分布）而高于 1% 临界值 -2.54。本例的 p 值可按下式计算：

$$p\text{ 值} = \mathrm{P}(T_{19} < -2.13) \tag{C.41}$$

式中，T_{19} 为自由度为 19 的 t 分布随机变量。此不等式与式（C.40）中的相反，这是因为它属于对立形式（C.33）而不是式（C.32），式（C.41）中的概率就是 t_{19} 分布中 -2.13 以右的面积（见图 C.8）。

733

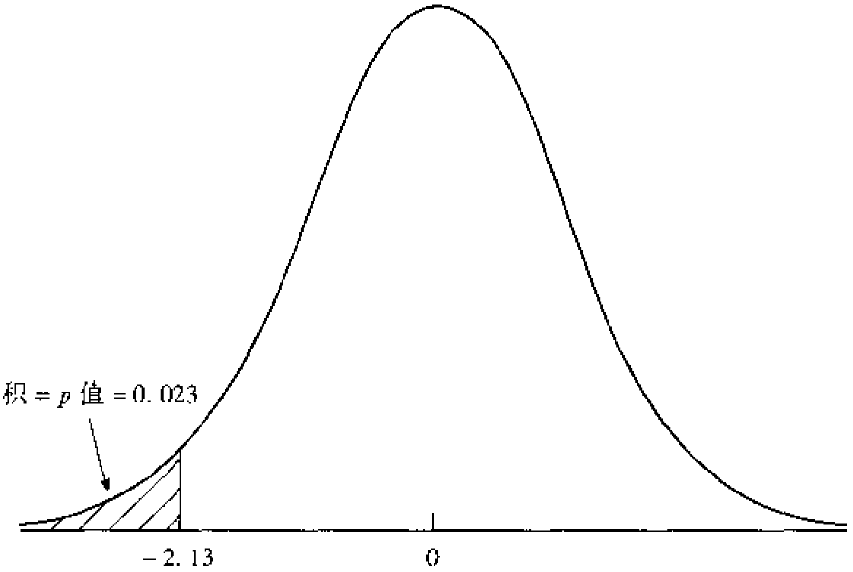


图 C.8　对于单侧对立 $\mu<0$，自由度为 19，$t=-2.13$ 时的 p 值

利用表 G.2，我们只知道 p 值介于 0.025～0.01 之间，但它比较接近 0.025（因为第 97.5 百分位数约为 2.09）。利用诸如 Stata 的统计软件包我们

能计算精确的 p 值。计算结果是约为 0.023，这对 H_0 的怀疑提供了合理证据。于是，我们有足够的证据在 2.5%（更不用说 5%）显著水平上拒绝培训津贴没有效果的虚拟假设。

计算双侧检验的 p 值，方法是类似的，但我们必须考虑双侧拒绝规则的性质。为了对总体均值作 t 检验，p 值计算如下：

$$P(|T_{n-1}|>|t|)=2P(T_{n-1}>|t|) \qquad (C.42)$$

式中，t 为检验统计量的值；T_{n-1} 为 t 随机变量。（对于大的 n，用标准正态随机变量代替 T_{n-1}。）于是，计算 t 统计量的绝对值，求 t_{n-1} 分布中此值以右的面积，再取其 2 倍。

734 对正非正态总体，精确的 p 值也许不容易求，但可通过同样的计算找到渐近的 p 值，这样的 p 值对大的样本是成立的。比方说，对 $n>120$，我们就可利用标准正态分布，表 G.1 是一个详细的、可以获得准确 p 值的表，但我们也可利用一个统计学或计量经济学程序。

例 C.7　员工雇用中的种族歧视

利用城市管理部门的配对数据（$n=241$），我们算得 $t=-4.29$。令 Z 为标准正态随机变量，则从实际角度看 P（$Z<-4.29$）就等于零。换句话说，在本例中，（渐近的）p 值几乎是零。这是拒绝 H_0 的强有力证据。

小结：怎样使用 p 值

(i) 选取检验统计量 T 并明确对立假设的性质，从而决定使用的拒绝规则是 $t>c$，$t<-c$，或者 $|t|>c$。

(ii) 以 t 统计量的观测值作为临界值，并计算该检验的相应显著水平。这就是 p 值。如果拒绝规则属于 $t>c$ 的形式，则 p 值 $=P(T>t)$。如果拒绝规则是 $t<-c$，则 p 值 $=P(T<t)$；如果拒绝规则是 $|t|>c$，则 p 值 $=P(|T|>|t|)$。

(iii) 如果显著水平 α 已经选定，那么，当 p 值 $<\alpha$ 时我们以 $100\cdot\alpha\%$ 水平拒绝 H_0；而当 p 值 $\geqslant\alpha$ 时我们未能按 $100\cdot\alpha\%$ 水平拒绝 H_0。也就是说，小的 p 值才会导致拒绝 H_0。

置信区间与假设检验的关系

因为构造置信区间和作假设检验都涉及概率的表述，所以想像它们之间有某种联系是自然的。事情确实如此。在置信区间构成之后，我们便可进行种种假设检验。

从性质上考虑，我们所讨论的置信区间都是双侧的。（在本教材中我们不需要构造单侧置信区间。）因此，置信区间可用于检验有双侧对立假设的虚拟假设。对于总体均值的检验，虚拟假设由式（C.31）给出，其对立假设为式

(C.34)。假定我们已构造 μ 的一个 95% 置信区间。于是，如果在 H_0 下，μ 的假设值 μ_0 不在该置信区间上，则我们以 5% 水平拒绝 $H_0: \mu = \mu_0$ 而选取 $H_1: \mu \neq \mu_0$。反之，如果 μ_0 落入该区间，则我们未能在 5% 水平上拒绝 H_0。应注意到，一旦置信区间被构成，对 μ_0 的任何值都可加以检验，并且由于置信区间包含不止一点，故可以有多个虚拟假设不被拒绝。

例 C.8　培训津贴与工人生产率

735 在霍尔泽等人的例子中，我们构造了废弃率平均变化 μ 的一个 95% 置信区间为 $[-2.28, -0.02]$。由于此区间不包含零，我们就能在 5% 水平上拒绝 $H_0: \mu = 0$ 而选取 $H_1: \mu \neq 0$。该置信区间还告诉我们，我们不能在 5% 水平上拒绝 $H_0: \mu = -2$。事实上，给定这个置信区间，存在一个我们所不能拒绝的虚拟假设的连续统。

实际显著性与统计显著性的对比

在我们至今讨论过的例子里，我们给出了表征总体参数的三种方式：点估计、置信区间和假设检验。这些工具对于了解总体参数都是同等重要的。目前存在一种可以理解的倾向，要求学生专心关注置信区间和假设检验，因为我们能把置信和显著水平的概念同它们联系起来。然而，不管做什么研究，我们还必须解释点估计值的大小或尺度。

统计显著性依赖于 t 统计量的大小，而不仅依赖于 $\bar{y}$ 的大小。为了检验 $H_0: \mu = 0$，我们需要 $t = \bar{y}/\mathrm{se}(\bar{y})$。因此统计显著性依赖于 $\bar{y}$ 与它的标准误之比。t 统计量大也许是因为 $\bar{y}$ 大，也许是因为 $\mathrm{se}(\bar{y})$ 小。

例 C.9　高速公路宽度对交通时间的影响

设 Y 为乘车到大都市上班的人在公路加宽前后所花时间的改变量，以分钟计。假定 $Y \sim \mathrm{Normal}\ (\mu,\ \sigma^2)$。表示加宽公路没有减少平均交通时间的虚拟假设是 H_0：$\mu = 0$；表示减少平均交通时间的对立假设是 $H_1: \mu < 0$。假定为了确定公路的加宽计划是否有效，我们抽取了一个大小为 $n = 300$ 的上下班时间的随机样本。经计算，交通时间的平均变化为 $\bar{y} = -3.6$，而样本标准差为 $s = 18.7$；因而 $\mathrm{se}(\bar{y}) = 18.7/\sqrt{300} \approx 1.08$。$t$ 统计量是 $t = -3.6/1.08 \approx -3.33$，这是统计上非常显著的；其 p 值基本上等于零。所以我们的结论是公路加宽对平均上下班交通时间有统计上的显著效果。

如果研究的报告仅限于假设检验的结果，那它就是误导的。只报告统计显著性将掩盖一个事实：所估计的平均交通时间的减少不过是一个小得可怜的数——3.6 分钟，面对事实，我们应该连同显著性检验一起，同时报告点估计值 3.6。736

虽然 t 统计量的大小和符号决定了统计上的显著性，但点估计 $\bar{y}$ 却决定着我们所称谓的**实际显著性**（practical significance）或重要性。一个估计值可以是统计上显著的而无须特别大。我们应经常连同点估计值的统计显著性一起讨论它的实际显著性。这一主题将在本书中反复出现。

当我们处理大样本时，会常常发现统计上显著的点估计值并不是实际上显著的。为了讨论清楚为什么会发生这种情况，下面的定义是有用的。

检验的一致性

只要 H_1 是真实的，一个一致性检验就会随着样本大小的增大而以趋于 1 的概率拒绝 H_0。

检验是一致性的另一说法是，每逢 H_1 是正确时，该检验的功效将随样本大小趋于无穷而有越来越接近于 1 的功效。本书中所遇到的检验，全都有这一性质。拿关于总体均值的假设检验来说，由于 Y 的方差随着样本大小的变大而收敛于零，便因此推知检验的一致性。检验 H_0：$\mu=0$ 的 t 统计量是 $T=\bar{Y}/(S/\sqrt{n})$。既然 $\text{plim}(\bar{Y})=\mu$ 且 $\text{plim}(S)=\sigma$，由此推知，如果 $\mu>0$（比方说），则 T 随着 $n\to\infty$ 而（以高的概率）变得越来越大。换句话说，不管 μ 多么接近于零，对于足够大的样本，我们能够拒绝 H_0 几乎是肯定无疑的。这里一点都不涉及 μ 从实际意义上看是否大。

C.7 关于符号的注释

在这里的概率与统计学复习中，以及在附录 B 中，我们细心地按照标准的惯例去标识随机变量、估计量和检验统计量。例如，我们用 W 表示一个估计量（随机变量）而用 w 表示一个特殊的估计值（随机变量 W 的结果）。区分一个估计量和一个估计值，以理解估计和假设检验中的种种概念是重要的。然而，由于一些模型有更大程度的复杂性，要做好这种区分很快就变成了计量经济学分析中的一种负担：模型会涉及很多随机变量和参数，为了忠实于概率与统计学中通用的惯例，就需要添加许多符号。

在正文中，我们用了计量经济学中广为采用的一种较简单的惯例。如果 θ 是总体参数，记号 $\hat{\theta}$（读"theta 帽"）就用来兼指 θ 的估计量和估计值。这种符号的好处，在于它用简单的方法把一个估计量和它想要估计的总体参数联系起来。这样，如果总体参数是 β，那么 $\hat{\beta}$ 就表示 β 的估计量或估计值；如果参数是 σ^2，$\hat{\sigma}^2$ 便是 σ^2 的估计量或估计值；等等。有时，我们讨论同一参数的两种估计量，这时需要用一个不同的符号，比如 $\tilde{\theta}$（读"theta 波"）。

放弃了概率与统计学中用以标识随机变量、估计量和检验统计量的惯例虽然加重了读者的负担，可是当你一旦把估计量和估计值的区别弄明白了，这也算不了什么负担。当我们在讨论 $\hat{\theta}$ 的统计性质——诸如推导它是否无偏

737 或一致——时，我们必然把 $\hat{\theta}$ 看成一个估计量。另一方面，如果写 $\hat{\theta}=1.73$，则我们明显地在指来自给定样本数据的一个点估计值。一旦你对概率与统计学有了良好的理解，由于使用 $\hat{\theta}$ 而造成模棱两可的混乱就能减少到最小限度。

▶ 小结

我们讨论了数理统计中为计量经济分析所高度依赖的一些问题。基本的一个概念是统计量，它不外是为了估计总体参数把数据组合起来的一个规则。我们讨论了估计量的种种性质，最重要的小样本性质是无偏性和有效性。后者有赖于在无偏估计量当中比较其方差。大样本性质关系到样本不断增大时得到的一系列估计量的情况，并且在计量经济学中经常要用到它们。凡是有用的估计量都是一致性的。中心极限定理的含义是，大多数估计量的抽样分布对大样本来说都是近似于正态的。

一个估计量的抽样分布可用于构造置信区间，在估计正态分布的均值和计算非正态情形的近似置信区间时，我们看到了这点。经典的假设检验要求设定一个虚拟假设、一个对立假设和一个显著水平，然后用一个检验统计量去和一个临界值相比而加以完成。另一方法是，计算一个 p 值，使得我们能够在任何一个显著水平上完成这个检验。

关键术语

对立假设
渐近正态性
偏误
中心极限定理（CLT）
置信区间
一致（性）估计量
一致（性）检验
临界值
估计值
估计量
假设检验
非一致的
区间估计量
大数定律（LLN）
检验的功效
实际显著性（重要性）
概率极限
p 值
随机样本
拒绝域
样本均值
样本相关系数
样本协方差
样本标准差
样本方差
抽样分布
抽样方差
显著（性）水平

最小二乘估计量
最大似然估计量
均方误（MSE）
矩法
最小方差无偏估计量
虚拟假设
单侧对立（假设）
单尾检验
总体
标准误
t 统计量
检验统计量
双侧对立（假设）
双尾检验
第Ⅰ类错误
第Ⅱ类错误
无偏性

习 题

738 **C.1** 令 Y_1、Y_2、Y_3 和 Y_4 为来自均值为 μ、方差为 σ^2 的总体的独立同分布随机变量；令 $\bar{Y} = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$ 为这 4 个随机变量的均值。

(i) 用 μ 和 σ^2 表示的 $\bar{Y}$ 的期望值和方差是什么？

(ii) 现在考虑的一个不同的估计量：

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4$$

这是 Y_i 的加权平均的一个例子。证明 W 也是 μ 的一个无偏估计量；求 W 的方差。

(iii) 根据你对(i)和(ii)的答案，你认为 μ 的哪个估计量好些，$\bar{Y}$ 还是 W？

(iv) 现在考虑由下式定义的 μ 的更一般的估计量：

$$W_a = a_1Y_1 + a_2Y_2 + a_3Y_3 + a_4Y_4$$

式中，a_i 为常数。为使 W_a 成为 μ 的一个无偏估计量，需要对这些 a_i 加上什么条件？

(v) 计算（iv）中的估计量的方差。

C.2 这是习题 C.1 的一个更一般的形式。令 $Y_1, Y_2, \cdots, Y_n$ 为有相同均值 μ 和相同方差 σ^2 的两两不相关随机变量；令 Y 为样本均值。

(i) 定义 μ 的线性估计量为

$$W_a = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$$

式中，a_i 为常数。为使 W_a 成为 μ 的一个无偏估计量，需要对 a_i 加上什么约束？

(ii) 求 $\mathrm{Var}(W_a)$。

(iii) 对任何数 $a_1, a_2, \cdots, a_n$，以下不等式恒成立：$(a_1 + a_2 + \cdots + a_n)^2/n \leqslant a_1^2 + a_2^2 + \cdots + a_n^2$。利用此不等式连同（i）和（ii），证明只要 W_a 是无偏的就有 $\mathrm{Var}(W_a) \geqslant \mathrm{Var}(Y)$，因而 Y 是最优线性无偏估计量。[提示：当

a_i 满足（i）所施加的约束时，不等式会变成什么？]

C.3 令 $\bar{Y}$ 为有均值 μ 和方差 σ^2 的随机样本的样本均值。考虑 μ 的两个不同的估计量：$W_1=[(n-1)/n]Y$ 和 $W_2=Y/2$。

（i）说明 W_1 和 W_2 都是 μ 的偏误估计量并求出此偏误。随着 $n\to\infty$，这些偏误会变成什么情形？试评论随着 n 变大，两估计量的偏误有何重要的差异？

（ii）求 W_1 和 W_2 的概率极限。[提示：利用性质 PLIM.1 和 PLIM.2；对于 W_1，注意 $\text{plim}[(n-1)/n]=1$。]哪一个估计量是一致的？

（iii）求 $\text{Var}(W_1)$ 和 $\text{Var}(W_2)$。

（iv）论证当 μ "接近"于零时 W_1 是比 Y 更好的一个估计量。（兼考虑偏误和方差。）

739 **C.4** 对于正的随机变量 X 和 Y，假定在给定 X 下，Y 的期望值是

$$\mathrm{E}(X|Y)=\theta X$$

未知的参数 θ 表示了 Y 的期望值是怎样随 X 而变的。

（i）定义随机变量 $Z=Y/X$。证明 $\mathrm{E}(Z)=\theta$。[提示：利用性质 CE.2 连同重期望律——性质 CE.4。具体地说，先证明 $\mathrm{E}(Z|X)$，再用性质 CE.4。]

（ii）利用（i）证明估计量 $W=n^{-1}\sum_{i=1}^{n}(Y_i/X_i)$ 对 W 是无偏的，其中 $\{(X_i,Y_i): i=1,2,\cdots,n\}$ 是一随机样本。

（iii）下表列出艾奥瓦州几个县的谷物产量数据。USDA 根据卫星照片预报了每县的谷物公顷数。研究者在卫星照片上辨认用于种植谷物的格数［以便除掉用于种植黄豆（比方说）或没有任何种植的格数］，并用这些格数去预报实际的（用于种植谷物的）公顷数。为了推导出一个可通用于各县的预报方程，USDA 调查了一些选定县里的农业主，以获得每公顷的谷物产量。令 Y_i = 第 i 县的谷物产量，并令 X_i = 第 i 县卫星照片中的谷物格数。一共对 8 个县作了 $n=17$ 次观测。试用这个样本来计算（ii）中所设计的对 θ 的估计量。

耕种地	谷物产量	种谷物格数
1	165.76	374
2	96.32	209
3	76.08	253
4	185.35	432
5	116.43	367
6	162.08	361
7	152.04	288

740

8	161.75	369
9	92.88	206
10	149.94	316
11	64.75	145
12	127.07	355
13	133.55	295
14	77.70	223
15	206.39	459
16	108.33	290
17	118.17	307

C.5 令 Y 为一Bernoulli（θ）（$0<\theta<1$）随机变量。假定我们感兴趣于估计机会比率 $\gamma=\theta/(1-\theta)$，即成功概率与失败概率之比。给定一个随机样本 $\{Y_1,\cdots,Y_n\}$，我们知道 θ 的一个无偏且一致估计量是 $\bar{Y}$，即 n 次试验中成功次数占的比例。γ 的一个自然估计量是 $G=\{\bar{Y}/(1-\bar{Y})\}$，即成功的比例与失败的比例之比。

（i）为什么 G 不是 γ 的无偏估计量？

（ii）利用 PLIM.2（iii），说明 G 是 γ 的一致估计量。

C.6 州长雇用你去研究一项饮料税是否会减少你所在州的平均饮料消费。你有能力对一个随机抽取的个人样本在课税之前和之后的年份里获得饮料消费的差异（以盎司计算）。对从总体随机抽取的第 i 个人，用 Y_i 表示他在饮料消费上的变化。把这些 Y_i 看成来自一个 Normal（μ，σ^2）总体的随机样本。

（i）虚拟假设是平均饮料消费没有变化。试用 μ 正式把它表述出来。

（ii）对立假设是饮料消费有所减少。用 μ 表示这个对立。

（iii）现在假定你的样本大小是 $n=900$，并且得到估计值 $\bar{y}=-32.8$ 和 $s=466.4$。计算 H_1 以作为对立的用以检验 H_0 的 t 统计量；求该检验的 p 值。（因为有了一个大的样本大小，尽可使用表 G.1 所列的标准正态分布。）你会在 5%水平上拒绝 H_0 吗？在 1%水平上又如何？

（iv）你会说所估计的消费下降在尺度上算是大的吗？评论这个估计值的实际和统计显著性。

741 （v）为了推测从税收变化到饮料消费之间的因果关系，在你的分析中，你对这两年期间决定饮料消费的因素方面作了什么隐含的假定？

C.7 面包房的新经理声称，现在的工人生产率高于旧经理时代，因此，工资都“普遍增加”了，令 W_i^b 为旧经理时代工人 i 的工资；W_i^a 为新经理时代

工人 i 的工资，差值为 $D_i = W_i^a - W_i^b$。假定 D_i 是来自 Normal(μ, σ^2) 分布的一个随机变量。

(i)利用下表中 15 个工人的数据构造 μ 的一个 95%置信区间。

(ii)正式表述平均工资没有变化这个虚拟假设，特别是在 H_0 之下，$E(D_i)$ 是什么？如果你被雇用来研究新经理声称的真实性，那么用 $\mu = E(D_i)$ 来表述的适当的对立假设是什么？

(iii)在 5%和 1%水平上，相对于所表述的对立假设，检验(ii)中的虚拟假设。

(iv)求(iii)中的检验的值。

工人	先前的工资	后来的工资
1	8.30	9.25
2	9.40	9.00
3	9.00	9.25
4	10.50	10.00
5	11.40	12.00
6	8.75	9.50
7	10.00	10.25
8	9.50	9.50
9	10.80	11.50
10	12.55	13.10
11	12.00	11.50
12	8.65	9.00
13	7.75	7.75
14	11.25	11.50
15	12.65	13.00

742

C.8 纽约时报（1/5/90）报导了全国篮球联赛（NBA）中前十名 3 分投篮手的 3 分投篮表现。下表总结了这些数据：

球　员	FGA-FGM
Mark Price	429 − 188
Trent Tucker	833 − 345
Dale Ellis	1 149 − 472
Craig Hodges	1 016 − 396
Danny Ainge	1 051 − 406
Byron Scott	676 − 260
Reggie Miller	416 − 159
Larry Bird	1 206 − 455
Jon Sundvold	440 − 166
Brian Taylor	417 − 157

说明：FGA 为投篮次数；FGM 为投中次数。

对给定的球员，一次特定投射的结果可模型化为一个贝努里（0—1）变量：令 Y_i 为投射的结果，如投中则 $Y_i = 1$；如不中则 $Y_i = 0$。令 θ 代表任一次 3 分投篮尝试成功的概率。θ 的自然估计量是 $\bar{Y} = FGM/FGA$。

（i）估计 Mark Price 的 θ。

（ii）求估计量 Y 的标准差，用 θ 和投篮的尝试次数 n 来表示这个标准差。

743 （iii）$(\bar{Y} - \theta)/\mathrm{se}(\bar{Y})$ 的渐近分布是标准正态的，其中 $\mathrm{se}(\bar{Y}) = \sqrt{\bar{Y}(1-\bar{Y})/n}$。利用这个事实，检验 Mark Price 的对立于 $H_1: \theta < 0.5$ 的虚拟假设 $H_0: \theta = 0.5$。

C.9　假定在一个不指名的国家，其军事独裁者举行了一次全民公决（是或否的信任投票）并声称他拥有 65% 的赞成票。一个人权组织怀疑有舞弊情况，并雇用你去检验独裁者所声称的真实性，你有在该国随机抽取 200 名投票人的足够经费。

（i）令 X 代表取自全体投票人总体的 200 人的随机样本的赞成票数。如果事实上有全体投票人的 65% 赞成这位独裁者，那么 X 的期望值是什么？

（ii）X 的标准差是什么？仍假定在全民公决中投赞成票的人的真正分数是 0.65。

（iii）现在，你收集了一个 200 人的样本并发现有 115 人真的投了赞成票。且假定真的有 65% 的全体公民表示赞成，那么，引用中心极限定理作近似计算，从一个大小为 200 的随机样本得到 115 或更少赞成票的概率是多少？

（iv）你怎样对一些没有经过统计学训练的人解释（iii）中的有关数字结

果呢？

C.10 在罢工过早地结束了1994年的主要联盟棒球赛季之前，San Diego Padres队的格温，在419次出击中击中165次，得到了一个0.394的棒击均值（batting average）。格温是不是当年的一位潜在0.400棒击者呢？曾有过一阵议论。这个问题可以用格温在某一次出击中击中的概率（把它叫做 θ）来重新表述。令 Y_i 为Bernoulli（θ）标示变量，如果格温在第 i 次出击时击中则取值1，否则取值零。于是 $Y_1, Y_2, \cdots, Y_n$ 构成一个来自Bernoulli（θ）分布的随机样本，其中 θ 是成功的概率，并且 $n=419$。

格温的棒击均值是我们对 θ 的最好估计，它也就是成功的比例：$\bar{y}=0.349$，利用 $\mathrm{se}(\bar{y})=\sqrt{\bar{y}(1-\bar{y})/n}$ 这一事实，再利用标准正态分布，构造 θ 的一个近似95%置信区间。你认为是否有强有力的证据足以拒绝格温是一位潜在的0.400棒击者？作出解释。

附录 D　矩阵代数概述

744　本附录概括了矩阵代数的一些概念(包括概率代数),以供研究使用附录E中矩阵表达的多元线性回归模型之需。这些内容在正文中都没有用到过。

D.1　基本定义

定义 D.1　(矩阵)

一个**矩阵**(matrix)就是一个矩形数组。更准确地讲,一个 $m\times n$ 矩阵就有 m 行和 n 列。正整数 m 被称为**行维数**,n 被称为**列维数**。

我们用大写黑体字母表示矩阵。一般可以将一个 $m\times n$ 矩阵写成

$$\mathbf{A}=[a_{ij}]=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

式中,a_{ij} 表示第 i 行和第 j 列的元素。比如 a_{25} 表示 $\mathbf{A}$ 中第 2 行和第 5 列的数字。一个具体的 2×3 矩阵的例子是

$$A=\begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix} \qquad (D.1)$$

式中，$a_{13}=7$。简略形式 $A=[a_{ij}]$ 通常用于定义矩阵。

定义 D.2 （方阵）

方阵（square matrix）具有相同的行数和列数。一个方阵的维数就是其行数和列数。

745 **定义 D.3 （向量）**

(i) 一个 $1\times m$ 的矩阵被称为一个（m 维）**行向量**（row vector），并可记为

$$\boldsymbol{x}\equiv(x_1,x_2,\cdots,x_m)$$

(ii) 一个 $n\times 1$ 的矩阵被称为一个**列向量**（column vector），并可记为

$$\boldsymbol{y}\equiv\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

定义 D.4 （对角矩阵）

当一个方阵 A 的非对角元素都为零（即对所有的 $i\neq j$ 都有 $a_{ij}=0$）时，它就是一个**对角矩阵**（diagonal matrix）。我们总能将一个对角矩阵写成

$$A=\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{mn} \end{bmatrix}$$

定义 D.5 （单位矩阵和零矩阵）

(i) 用 $\boldsymbol{I}$（或为了强调维数而用 $\boldsymbol{I}_n$）表示的 $n\times n$ **单位矩阵**（identity matrix）就是对角位置都是 1 而其他位置都是 0 的对角阵：

$$\boldsymbol{I}\equiv\boldsymbol{I}_n\equiv\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

(ii) 一个用 $\boldsymbol{0}$ 表示的 $m\times n$ **零矩阵**（zero matrix）就是所有元素都为零的 $m\times n$ 矩阵。它不一定是方阵。

D.2 矩阵运算

矩阵加法

两个都是 $m\times n$ 维的矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 可通过对应元素相加而相加：$\boldsymbol{A}+\boldsymbol{B}=$

$[a_{ij}]+[b_{ij}]$。更准确地，有

$$A+B=\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

746 比如

$$\begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}+\begin{bmatrix} 1 & 0 & -4 \\ 4 & 2 & 3 \end{bmatrix}=\begin{bmatrix} 3 & -1 & 3 \\ 0 & 7 & 3 \end{bmatrix}$$

不同维数的矩阵不能相加。

数乘

给定任意一个实数 γ（常被称为一个数量），**数乘**（scalar multiplication）被定义为 $\gamma A\equiv[\gamma a_{ij}]$，或

$$\gamma A=\begin{bmatrix} \gamma a_{11} & \gamma a_{12} & \cdots & \gamma a_{1n} \\ \gamma a_{21} & \gamma a_{22} & \cdots & \gamma a_{2n} \\ \vdots & \vdots & & \vdots \\ \gamma a_{m1} & \gamma a_{m2} & \cdots & \gamma a_{mn} \end{bmatrix}$$

比如，若 $\gamma=2$，且 A 是等式（D.1）中的矩阵，则

$$\gamma A=\begin{bmatrix} 4 & -2 & 14 \\ -8 & 10 & 0 \end{bmatrix}$$

矩阵乘法

为了使矩阵 A 乘以矩阵 B 以得到 AB，A 的列维数和 B 的行维数必须相同。因此，令 A 为一个 $m\times n$ 矩阵，而 B 为一个 $n\times p$ 矩阵。于是**矩阵乘法**（matrix multiplication）被定义为

$$AB=\left[\sum_{k=1}^{n}a_{ik}b_{kj}\right]$$

换句话说，新矩阵 AB 的第（i，j）个元素，等于 A 中第 i 行的每个元素与 B 中第 j 列对应元素的乘积之和。如下简图可以使这个过程更一目了然：

$$\begin{array}{cccc} & \mathbf{A} & \mathbf{B} & \mathbf{AB} \\ \text{第 } i \text{ 行} & \begin{bmatrix} & & & & \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ & & & & \end{bmatrix} & \begin{bmatrix} b_{1j} \\ b_{2j} \\ b_{3j} \\ \vdots \\ b_{nj} \end{bmatrix} = & \begin{bmatrix} \\ \sum_{k=1}^{n} a_{ik}b_{kj} \\ \\ \end{bmatrix} \\ & & \text{第 } j \text{ 列} & \text{第 } (i,\ j) \text{ 个元素} \end{array}$$

747 其中根据附录 A 中求和运算的定义，有

$$\sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

如

$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}$$

我们也可以将一个矩阵与一个向量相乘。如果 $\mathbf{A}$ 是一个 $n \times m$ 矩阵，而 $\mathbf{y}$ 是一个 $m \times 1$ 的向量，那么 $\mathbf{Ay}$ 就是一个 $n \times 1$ 的向量。如果 $\mathbf{x}$ 是一个 $1 \times n$ 的向量，那么 $\mathbf{xA}$ 就是一个 $1 \times m$ 的向量。

矩阵加法、数乘和矩阵乘法可以用各种方式组合，而且这些运算还满足几个熟悉的基本数值运算规则。在如下性质表中，$\mathbf{A}$，$\mathbf{B}$ 和 $\mathbf{C}$ 都具有运算所需要的适当的维数，而 α 和 β 则是实数。这些性质中的大多数都很容易从定义得到说明。

矩阵乘法的性质：

(1)$(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$；

(2)$\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$；

(3)$(\alpha\beta)\mathbf{A} = \alpha(\beta\mathbf{A})$；

(4)$\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B}$；

(5)$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$；

(6)$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$；

(7)$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$；

(8)$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$；

(9)$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$；

(10)$\mathbf{IA} = \mathbf{AI} = \mathbf{A}$；

(11)$\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$；

(12)$\mathbf{A} - \mathbf{A} = \mathbf{0}$；

(13)$\mathbf{A0} = \mathbf{0A} = \mathbf{0}$；

(14)即使 $\mathbf{AB}$ 和 $\mathbf{BA}$ 都有定义，仍然会 $\mathbf{AB} \neq \mathbf{BA}$。

对最后一个性质值得进一步评论。如果 **A** 是一个 $n\times m$ 矩阵，而 **B** 是一个 $m\times p$ 矩阵，那么 **AB** 就有定义；而 **BA** 只有在 $n=p$（**A** 的行维数等于 **B** 的列维数）时有定义。如果 **A** 是一个 $m\times n$ 矩阵，而 **B** 是一个 $n\times m$ 矩阵，那么 **AB** 和 **BA** 就都有定义；但除非 **A** 和 **B** 都是方阵，否则它们具有不同的维数；即便 **A** 和 **B** 都是方阵，除非在特殊情形下，否则 $\boldsymbol{AB}\neq\boldsymbol{BA}$ 仍成立。

转置

定义 D.6　（转置）

令 $\boldsymbol{A}=[a_{ij}]$ 表示一个 $m\times n$ 矩阵。用 $\boldsymbol{A}'$（读作 **A** 撇）表示 **A** 的转置，是将 **A** 的行和列互换后得到的 $n\times m$ 矩阵。我们可以把它写成 $\boldsymbol{A}'\equiv[a_{ji}]$。比如

$$\boldsymbol{A}=\begin{bmatrix}2 & -1 & 7\\ -4 & 5 & 0\end{bmatrix},\qquad \boldsymbol{A}'=\begin{bmatrix}2 & -4\\ 1 & 5\\ 7 & 0\end{bmatrix}$$

转置的性质：

748　(1) $(\boldsymbol{A}')'=\boldsymbol{A}$；

(2) $(\alpha\boldsymbol{A})'=\alpha\boldsymbol{A}'$，$\alpha$ 为任意数；

(3) $(A+B)'=A'+B'$；

(4) $(AB)'=B'A'$，A 和 B 分别是 $m\times n$ 和 $n\times k$ 矩阵；

(5) $\boldsymbol{x}'\boldsymbol{x}=\sum_{i=1}^{n}x_i^2$，$\boldsymbol{x}$ 是一个 $n\times 1$ 向量；

(6) 如果 **A** 是一个各行分别由 $1\times k$ 的向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$ 给出的 $n\times k$ 矩阵，所以可以写成

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_n\end{bmatrix}$$

于是　$\boldsymbol{A}'=(\boldsymbol{a}_1'\ \boldsymbol{a}_2'\ \cdots\ \boldsymbol{a}_n')$

定义 D.7　（对称矩阵）

一个方阵是一个对称矩阵的充分必要条件是 $\boldsymbol{A}'=\boldsymbol{A}$。

如果 **X** 是任何一个 $n\times k$ 矩阵，那么 $\boldsymbol{X}'\boldsymbol{X}$ 总有定义并是一个对称矩阵，通过应用转置的第一和第四条性质即可看出（参见习题 D.3）。

分块矩阵的乘法

令 **A** 表示一个行由 $1\times k$ 向量 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_n$ 给出的一个 $n\times k$ 矩阵，令 **B** 表示一个行由 $1\times m$ 向量 $\boldsymbol{b}_1$，$\boldsymbol{b}_2$，…，$\boldsymbol{b}_n$ 给出的一个 $n\times m$ 矩阵：

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_n\end{bmatrix},\qquad \boldsymbol{B}=\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\\ \vdots\\ \boldsymbol{b}_n\end{bmatrix}$$

那么

$$\boldsymbol{A}'\boldsymbol{B}=\sum_{i=1}^{n}\boldsymbol{a}'_i\boldsymbol{b}_i$$

式中，$\boldsymbol{a}'_i\boldsymbol{b}_i$ 对每个 i，都是一个 $k\times m$ 矩阵。因此，$\boldsymbol{A}'\boldsymbol{B}$ 可写成 n 个 $k\times m$ 矩阵之和。作为一个特殊情形，我们有

$$\boldsymbol{A}'\boldsymbol{A}=\sum_{i=1}^{n}\boldsymbol{a}'_i\boldsymbol{a}_i$$

式中，$\boldsymbol{a}'_i\boldsymbol{a}_i$ 对所有的 i 都是一个 $k\times k$ 矩阵。

迹

一个矩阵的迹是只对方阵定义的一个很简单的运算。

定义 D.8 （迹）

对任何一个 $n\times n$ 矩阵 $\boldsymbol{A}$，用 $\mathrm{tr}(\boldsymbol{A})$ 表示的**矩阵 $\boldsymbol{A}$ 的迹**（trace of a matrix $\boldsymbol{A}$）就是其主对角线元素之和。从数学上看，即

749

$$\mathrm{tr}(\boldsymbol{A})=\sum_{i=1}^{n}a_{ii}$$

迹的性质：

(1) $\mathrm{tr}(\boldsymbol{I}_n)=n$；

(2) $\mathrm{tr}(\boldsymbol{A}')=\mathrm{tr}(\boldsymbol{A})$；

(3) $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})=\mathrm{tr}(\boldsymbol{A})+\mathrm{tr}(\boldsymbol{B})$；

(4) 对任意数量 α，都有 $\mathrm{tr}(\alpha\boldsymbol{A})=\alpha\,\mathrm{tr}(\boldsymbol{A})$；

(5) $\mathrm{tr}(\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{BA})$，$\boldsymbol{A}$ 是 $m\times n$ 矩阵，而 $\boldsymbol{B}$ 是 $n\times m$ 矩阵。

逆

对方阵而言，逆矩阵是一个很重要的概念。

定义 D.9 （逆）

对一个 $n\times n$ 矩阵 $\boldsymbol{A}$，如果 $\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{I}_n$，则 $\boldsymbol{A}^{-1}$ 表示矩阵 $\boldsymbol{A}$ 的逆（inverse）。在这种情况下，$\boldsymbol{A}$ 就是可逆的或非奇异的。否则，它就是不可逆的

或奇异的。

逆的性质：

(1)如果逆存在，它是惟一的；

(2)如果 $\alpha \neq 0$，且 $\boldsymbol{A}$ 是可逆的，则 $(\alpha\boldsymbol{A})^{-1} = (1/\alpha)\boldsymbol{A}^{-1}$；

(3)如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是 $n \times n$ 可逆矩阵，则 $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$；

(4)$(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$。

我们不讨论计算一个矩阵的逆的步骤。任何一本矩阵代数的教科书都会有这种计算的详细例子。

D.3 线性独立与矩阵的秩

对一组具有相同维数的向量而言，知道其中一个向量是否可表示成其余向量的线性组合很重要。

定义 D.10 （线性独立）

令 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r\}$ 是一组向量。当且仅当

$$\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \cdots + \alpha_r\boldsymbol{x}_r = \boldsymbol{0} \qquad \text{(D.2)}$$

$\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$ 时，它们才是**线性独立的向量**（linearly independent vector）。如果式（D.2）对一组不全为零的系数成立，那么 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r\}$ 就是**线性相关**的。

$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r\}$ 线性相关的说法等于说，其中至少有一个向量可写成其他向量的线性组合。

定义 D.11 （秩）

(i) 令 $\boldsymbol{A}$ 是一个 $n \times m$ 矩阵。用 rank（$\boldsymbol{A}$）表示的**矩阵 $\boldsymbol{A}$ 的秩**（rank of matrix $\boldsymbol{A}$）就是 $\boldsymbol{A}$ 线性独立的最大列数。

(ii) 如果 $\boldsymbol{A}$ 是一个 $n \times m$ 矩阵，且 rank（$\boldsymbol{A}$）$= m$，那么 $\boldsymbol{A}$ 就具有**列满秩**（full column rank）。

750 如果 $\boldsymbol{A}$ 是一个 $n \times m$ 矩阵，那么它的秩最大为 m。如果一个矩阵的列构成一个线性独立集，那么它就是列满秩的。比如 3×2 矩阵

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 0 & 0 \end{bmatrix}$$

的秩最大为 2。实际上它的秩是 1，因为第二列是第一列的 2 倍。

秩的性质：

(1) rank（$\boldsymbol{A}'$）$=$ rank（$\boldsymbol{A}$）；

(2) 如果 $\boldsymbol{A}$ 是一个 $n \times k$ 矩阵，则 rank（$\boldsymbol{A}$）$\leqslant \min(n, k)$；

(3) 如果 $\boldsymbol{A}$ 是一个 $k \times k$ 矩阵，且 rank（$\boldsymbol{A}$）$= k$，则 $\boldsymbol{A}$ 是满秩的。

D.4 二次型与正定矩阵

定义 D.12 （二次型）

令 $\boldsymbol{A}$ 为 $n\times n$ 对称矩阵。与矩阵 $\boldsymbol{A}$ 相关的二次型，就是对所有 $n\times 1$ 向量 $\boldsymbol{x}$ 定义的实值函数：

$$f(\boldsymbol{x})=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}=\sum_{i=1}^{n}a_{ii}x_i^2+2\sum_{i=1}^{n}\sum_{j>i}a_{ij}x_ix_j$$

定义 D.13 （正定和半正定）

(i) 对于一个对称矩阵 $\boldsymbol{A}$，如果对除 $\boldsymbol{x}=\boldsymbol{0}$ 外的所有 $n\times 1$ 向量 $\boldsymbol{x}$，都有 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}>\boldsymbol{0}$，那么我们就说 $\boldsymbol{A}$ 是**正定的**（positive definite，p.d.）。

(ii) 对于一个对称矩阵 $\boldsymbol{A}$，如果对所有 $n\times 1$ 向量 $\boldsymbol{x}$，都有 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}\geqslant 0$，那么我们就说 $\boldsymbol{A}$ 是**半正定的**（positive semi-definite，p.s.d.）。

如果一个矩阵是正定或半正定的，那它就自动被假定为对称的。

正定和半正定矩阵的性质：

(1) 正定矩阵的主对角元素都严格为正，而半正定矩阵的主对角元素都非负；

(2) 如果 $\boldsymbol{A}$ 是正定矩阵，则 $\boldsymbol{A}^{-1}$ 存在并正定；

(3) 如果 $\boldsymbol{X}$ 是一个 $n\times k$ 矩阵，则 $\boldsymbol{X}'\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}'$ 都是半正定的；

(4) 如果 $\boldsymbol{X}$ 是一个 $n\times k$ 矩阵，且 rank（$\boldsymbol{X}$）$=k$，则 $\boldsymbol{X}'\boldsymbol{X}$ 是正定（因此也是满秩）的。

D.5 幂等矩阵

定义 D.14 （幂等矩阵）

令 $\boldsymbol{A}$ 为 $n\times n$ 对称矩阵。当且仅当 $\boldsymbol{A}\boldsymbol{A}=\boldsymbol{A}$ 时，我们称 $\boldsymbol{A}$ 为一个**幂等矩阵**（idempotent matrix）。

751 如

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&1\end{bmatrix}$$

就是一个幂等矩阵，直接相乘就可验证。

幂等矩阵的性质：

令 $\boldsymbol{A}$ 为 $n\times n$ 幂等矩阵。

(1) rank（$\boldsymbol{A}$）$=$ tr（$\boldsymbol{A}$）；

（2） $\boldsymbol{A}$ 是半正定矩阵。

我们可以构造一些很一般的幂等矩阵。令 $\boldsymbol{X}$ 表示一个 $\mathrm{rank}(\boldsymbol{X})=k$ 的 $n\times k$ 矩阵。定义

$$\boldsymbol{P}\equiv\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$
$$\boldsymbol{M}\equiv\boldsymbol{I}_n-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'=\boldsymbol{I}_n-\boldsymbol{P}$$

于是 $\boldsymbol{P}$ 和 $\boldsymbol{M}$ 是对称的幂等矩阵，且 $\mathrm{rank}(\boldsymbol{P})=k$ 和 $\mathrm{rank}(\boldsymbol{M})=n-k$。这些秩很容易通过利用性质(1)得到：$\mathrm{tr}(\boldsymbol{P})=\mathrm{tr}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}]$（根据迹的性质 5）$=\mathrm{tr}(\boldsymbol{I}_k)=k$（根据迹的性质 1）。接下来很容易得到 $\mathrm{tr}(\boldsymbol{M})=\mathrm{tr}(\boldsymbol{I}_n)-\mathrm{tr}(\boldsymbol{P})=n-k$。

D.6 线性形式和二次型的微分

对于一个给定的 $n\times 1$ 向量 $\boldsymbol{a}$，考虑对所有 $n\times 1$ 向量 $\boldsymbol{x}$ 由 $f(\boldsymbol{x})=\boldsymbol{a}'\boldsymbol{x}$ 定义的线性函数。f 对 $\boldsymbol{x}$ 的导数是 $1\times n$ 阶偏导数向量，也就是

$$\partial f(\boldsymbol{x})/\partial\boldsymbol{x}=\boldsymbol{a}'$$

对于一个 $n\times n$ 对称矩阵 $\boldsymbol{A}$，定义二次型 $g(\boldsymbol{x})=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$，于是 $\partial g(\boldsymbol{x})/\partial\boldsymbol{x}=2\boldsymbol{x}'\boldsymbol{A}$ 是一个 $1\times n$ 向量。

D.7 随机向量的矩和分布

为了利用矩阵推导 OLS 估计量的期望值和方差，我们需要定义一个**随机向量**(random vector)的期望值和方差。

顾名思义，一个随机向量无非就是随机变量的一个向量。我们还需要定义多元正态分布。这些概念无非就是对附录 B 中所讨论概念的推广。

期望值

752 **定义 D.15 （期望值）**

(i)如果 $\boldsymbol{y}$ 是一个 $n\times 1$ 随机向量，用 $\mathrm{E}(\boldsymbol{y})$ 表示的 $\boldsymbol{y}$ 的**期望值**(expected value)就是期望值向量：

$$\mathrm{E}(\boldsymbol{y})=[\mathrm{E}(y_1),\mathrm{E}(y_2),\cdots,\mathrm{E}(y_n)]'$$

(ii)如果 $\boldsymbol{Z}$ 是 $n\times m$ 随机矩阵，则 $\mathrm{E}(\boldsymbol{Z})$ 是期望值的 $n\times m$ 矩阵：

$$\mathrm{E}(\boldsymbol{Z})=[\mathrm{E}(z_{ij})]$$

期望值的性质：

(1)如果 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，而 $\boldsymbol{b}$ 是一个 $n\times 1$ 向量，且二者都是非随机的，那么 $\mathrm{E}(\boldsymbol{Ay}+\boldsymbol{b})=\boldsymbol{A}\mathrm{E}(\boldsymbol{y})+\boldsymbol{b}$；

(2)如果 $\boldsymbol{A}$ 是 $p\times n$ 矩阵，而 $\boldsymbol{B}$ 是 $m\times k$ 矩阵，且二者都是非随机的，那么 $\mathrm{E}(\boldsymbol{AZB})=\boldsymbol{A}\mathrm{E}(\boldsymbol{Z})\boldsymbol{B}$。

方差—协方差矩阵

定义 D.16 （方差—协方差矩阵）

如果 $\boldsymbol{y}$ 是一个 $n\times 1$ 随机向量，用 $\mathrm{Var}(\boldsymbol{y})$ 表示的 $\boldsymbol{y}$ 的**方差—协方差矩阵**（variance-covariance matrix）就定义为

$$\mathrm{Var}(\boldsymbol{y})=\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n}\\ \vdots & \vdots & & \vdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2\end{bmatrix}$$

式中，$\sigma_j^2=\mathrm{Var}(y_j)$ 和 $\sigma_{ij}=\mathrm{Cov}(y_i,y_j)$。换句话说，方差—协方差矩阵的主对角线上依次是 $\boldsymbol{y}$ 的每个元素的方差，而主对角线之外则是协方差。由于 $\mathrm{Cov}(y_i,y_j)=\mathrm{Cov}(y_j,y_i)$，所以立即知道方差—协方差矩阵是对称的。

方差的性质：

(1)若 $\boldsymbol{a}$ 是一个 $n\times 1$ 非随机向量，则 $\mathrm{Var}(\boldsymbol{a}'\boldsymbol{y})=\boldsymbol{a}'[\mathrm{Var}(\boldsymbol{y})]\boldsymbol{a}\geqslant 0$；

(2)若 $\mathrm{Var}(\boldsymbol{a}'\boldsymbol{y})>0$ 对所有 $\boldsymbol{a}\neq\boldsymbol{0}$ 都成立，则 $\mathrm{Var}(\boldsymbol{y})$ 是正定的；

(3)$\mathrm{Var}(\boldsymbol{y})=\mathrm{E}[(\boldsymbol{y}-\boldsymbol{\mu})(\boldsymbol{y}-\boldsymbol{\mu})']$，其中 $\boldsymbol{\mu}=\mathrm{E}(\boldsymbol{y})$；

(4)若 $\boldsymbol{y}$ 的各个元素都不相关，则 $\mathrm{Var}(\boldsymbol{y})$ 是一个对角阵。而且若对 $j=1,2,\cdots,n$ 都有 $\mathrm{Var}(y_j)=\sigma^2$，则 $\mathrm{Var}(\boldsymbol{y})=\sigma^2\boldsymbol{I}_n$；

(5)如果 $\boldsymbol{A}$ 是一个 $m\times n$ 非随机矩阵，而 $\boldsymbol{b}$ 是一个 $n\times 1$ 的非随机向量，那么 $\mathrm{Var}(\boldsymbol{Ay}+\boldsymbol{b})=\boldsymbol{A}[\mathrm{Var}(\boldsymbol{y})]\boldsymbol{A}'$。

多元正态分布

753 在附录B对一个随机变量的正态分布进行了某种程度的讨论。我们需要对正态分布扩展到随机向量的情形。由于不需要，所以我们将不给出概率分布函数的表达式。重要的是要知道，一个多元正态随机向量完全由其均值和方差—协方差矩阵所刻画。因此，如果 $\boldsymbol{y}$ 是一个均值为 $\boldsymbol{\mu}$ 和方差—协方差矩阵为 $\boldsymbol{\Sigma}$ 的 $n\times 1$ 多元正态随机向量，那么我们可以写成 $\boldsymbol{y}\sim\mathrm{Normal}(\boldsymbol{\mu},\boldsymbol{\Sigma})$。现在，我们来讨论**多元正态分布**（multivariate normal distribution）的几个有用的性质。

多元正态分布的性质：

（1）如果 $\boldsymbol{y}\sim\text{Normal}\,(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，那么 y 中的每个元素都是正态分布的。

（2）如果 $y\sim\text{Normal}\,(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，那么 y 中任意两个元素 y_i 和 y_j 相互独立的充分必要条件是它们不相关，即 $\sigma_{ij}=0$。

（3）如果 $\boldsymbol{y}\sim\text{Normal}\,(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，那么 $\boldsymbol{Ay}+\boldsymbol{b}\sim\text{Normal}\,(\boldsymbol{A\mu}+\boldsymbol{b},\ \boldsymbol{A\Sigma A}')$，其中 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 是非随机的。

（4）如果 $\boldsymbol{y}\sim\text{Normal}\,(0,\ \boldsymbol{\Sigma})$，那么对于非随机矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，$\boldsymbol{Ay}$ 和 $\boldsymbol{By}$ 独立的充分必要条件是 $\boldsymbol{A\Sigma B}'=\boldsymbol{0}$。特别是，若 $\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}_n$，则 $\boldsymbol{AB}'=\boldsymbol{0}$ 是 $\boldsymbol{Ay}$ 和 $\boldsymbol{By}$ 独立的充分必要条件。

（5）如果 $\boldsymbol{y}\sim\text{Normal}\,(0,\ \sigma^2\boldsymbol{I}_n)$，$\boldsymbol{A}$ 是一个 $k\times n$ 非随机矩阵，而 $\boldsymbol{B}$ 是一个 $n\times n$ 对称、幂等矩阵，那么 $\boldsymbol{Ay}$ 和 $\boldsymbol{y}'\boldsymbol{By}$ 独立的充分必要条件是 $\boldsymbol{AB}=\boldsymbol{0}$。

（6）如果 $\boldsymbol{y}\sim\text{Normal}\,(0,\ \sigma^2\boldsymbol{I}_n)$，而 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 又都是对称的幂等矩阵，那么 $\boldsymbol{y}'\boldsymbol{Ay}$ 和 $\boldsymbol{y}'\boldsymbol{By}$ 独立的充分必要条件是 $\boldsymbol{AB}=\boldsymbol{0}$。

χ 平方分布

我们在附录 B 中定义 **χ 平方随机变量**（chi-square random variable）为独立的标准正态随机变量的平方和。用向量表示为，若 $\boldsymbol{u}\sim\text{Normal}\,(\boldsymbol{0},\ \boldsymbol{I}_n)$，则 $\boldsymbol{u}'\boldsymbol{u}\sim\chi^2_n$。

χ 平方分布的性质：

（1）若 $\boldsymbol{u}\sim\text{Normal}\,(\boldsymbol{0},\ \boldsymbol{I}_n)$，且 $\boldsymbol{A}$ 是一个 rank（$\boldsymbol{A}$）$=q$ 的对称幂等矩阵，则 $\boldsymbol{u}'\boldsymbol{Au}\sim\chi^2_q$。

（2）若 $\boldsymbol{u}\sim\text{Normal}\,(\boldsymbol{0},\ \boldsymbol{I}_n)$，且 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是满足 $\boldsymbol{AB}=\boldsymbol{0}$ 的 $n\times n$ 对称幂等矩阵，则 $\boldsymbol{u}'\boldsymbol{Au}$ 和 $\boldsymbol{u}'\boldsymbol{Bu}$ 是相互独立的随机变量。

t 分布

我们在附录 B 中也定义了 **t 分布**（t distribution）。我们现在增加一个重要的性质。

t 分布的性质：

若 $\boldsymbol{u}\sim\text{Normal}(\boldsymbol{0},\boldsymbol{I}_n)$，$\boldsymbol{c}$ 是一个 $n\times 1$ 非随机向量，$\boldsymbol{A}$ 是一个秩为 q 且 $\boldsymbol{Ac}=\boldsymbol{0}$ 的非随机 $n\times n$ 对称、幂等矩阵，则 $\{\boldsymbol{c}'\boldsymbol{u}/(\boldsymbol{c}'\boldsymbol{c})^{1/2}\}/(\boldsymbol{u}'\boldsymbol{Au})^{1/2}\sim t_q$。

F 分布

记得两个独立的 χ 平方随机变量分别除以其自由度而标准化后的比值就是一个 **F 随机变量**（F random variable）。

F 分布的性质：

若 $\boldsymbol{u} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{I}_n)$，且 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是满足 $\text{rank}(\boldsymbol{A}) = k_1$，$\text{rank}(\boldsymbol{B}) = k_2$ 和 $\boldsymbol{AB} = \boldsymbol{0}$ 的 $n \times n$ 非随机对称幂等矩阵，则 $(\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}/k_1)/(\boldsymbol{u}'\boldsymbol{B}\boldsymbol{u}/k_2) \sim F_{k_1,k_2}$。

▶ 小结

本附录包含了用矩阵研究经典线性模型所需要的浓缩的背景信息。尽管这里的材料能自圆其说，但主要还是为那些熟悉矩阵代数和多元统计的读者作一回顾，而且将在附录 E 中广泛使用。

关键术语

754 χ 平方随机变量
列向量
对角阵
期望值
F 随机变量
幂等矩阵
单位（恒等）矩阵
逆矩阵
线性独立向量
矩阵
矩阵乘法
多元正态分布
正定

半正定
二次型
随机向量
矩阵的秩
行向量
数（量）乘（法）
方阵
对称矩阵
t 分布
矩阵的迹
转置
方差—协方差矩阵
零矩阵

习 题

D.1 （i）利用如下定义计算 $\boldsymbol{AB}$：

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} 0 & 1 & 6 \\ 1 & 8 & 0 \\ 3 & 0 & 0 \end{bmatrix}$$

(ii) $\boldsymbol{BA}$ 存在吗?

D.2 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是 $n\times n$ 对角阵，证明 $\boldsymbol{AB}=\boldsymbol{BA}$。

D.3 令 $\boldsymbol{X}$ 表示任何一个 $n\times k$ 矩阵，证明 $\boldsymbol{X}'\boldsymbol{X}$ 是一个对称矩阵。

D.4 (i) 利用迹的性质证明 $\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})=\mathrm{tr}(\boldsymbol{AA}')$，对任何 $n\times m$ 矩阵 $\boldsymbol{A}$ 都成立。

(ii) 用 $\boldsymbol{A}=\begin{bmatrix}2 & 0 & -1\\ 0 & 3 & 0\end{bmatrix}$ 验证 $\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})=\mathrm{tr}(\boldsymbol{AA}')$。

D.5 (i) 利用逆的定义证明：若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是 $n\times n$ 非奇异矩阵，则

$$(\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

(ii) 若 $\boldsymbol{A}$, $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 都是 $n\times n$ 非奇异矩阵，用 $\boldsymbol{A}^{-1}$, $\boldsymbol{B}^{-1}$ 和 $\boldsymbol{C}^{-1}$ 表示 $(\boldsymbol{ABC})^{-1}$。

D.6 (i) 证明：若 $\boldsymbol{A}$ 是一个 $n\times n$ 对称正定矩阵，则 $\boldsymbol{A}$ 的主对角元素一定严格为正。

(ii) 写出一个主对角元素严格为正但不是正定的 2×2 对称矩阵。

D.7 令 $\boldsymbol{A}$ 表示一个 $n\times n$ 对称正定矩阵。证明：若 $\boldsymbol{P}$ 是任何一个 $n\times n$ 非奇异矩阵，则 $\boldsymbol{P}'\boldsymbol{AP}$ 都是正定的。

D.8 利用向量方差的性质 3 证明性质 5。

附录 E　矩阵形式的线性回归模型

755 本附录利用矩阵符号和矩阵代数(参见附录D中的概述),推导了多元线性回归模型中普通最小二乘估计的各个结论。这里给出的材料比正文中的内容深入得多。

E.1 模型与普通最小二乘估计

在整个附录E中,我们都用下标 t 表示观测次数,并用 n 表示样本容量。如下写出含有 k 个参数的多元线性回归模型是有益的：

$$y_t=\beta_1+\beta_2x_{t2}+\beta_3x_{t3}+\cdots+\beta_kx_{tk}+u_t,\quad t=1,2,\cdots,n \qquad (\text{E}.1)$$

式中，y_t 为第 t 次观测的因变量，而 x_{tj}（$j=2, 3, \cdots, k$）都是自变量。注意我们这里的标记习惯与正文中不同：我们称 β_1 为截距项，而用 $\beta_2, \cdots, \beta_k$ 表示斜率参数。这种重新标记并不重要，但简化了多元回归的矩阵表述方法。

对每个 t，定义一个 $1\times k$ 向量 $\mathbf{x}_t=(1, x_{t2}, \cdots, x_{tk})$，并令 $\boldsymbol{\beta}=(\beta_1, \beta_2, \cdots, \beta_k)'$ 表示所有参数的 $k\times 1$ 向量。于是我们就可以将式（E.1）写成

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta} + u_t, \quad t = 1,2,\cdots,n \tag{E.2}$$

[有些作者更喜欢将 $\boldsymbol{x}_t$ 定义为一个列向量，在那种情况下，式（E.2）中的 $\boldsymbol{x}_t$ 将由 $\boldsymbol{x}'_t$ 替代。但从数学上讲，把它定义为一个行向量更有意义。] 通过适当地定义数据向量和矩阵，我们可以把式（E.2）完全写成矩阵的形式。令 $\boldsymbol{y}$ 表示 y 的观测数据向量：$\boldsymbol{y}$ 的第 t 个元素是 y_t。令 $\boldsymbol{X}$ 表示解释变量的观测数据的 $n\times k$ 矩阵。换句话说，$\boldsymbol{X}$ 的第 t 行由向量 $\boldsymbol{x}_t$ 构成。换言之，$\boldsymbol{X}$ 的第 756（t，j）个元素无非就是 x_{tj}：

$$\underset{n\times k}{\boldsymbol{X}} \equiv \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}$$

最后，令 $\boldsymbol{u}$ 表示观测不到的误差项 $n\times 1$ 向量。于是，我们可以将所有 n 个观测的式（E.2）写成**矩阵形式**（matrix notation）：

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \tag{E.3}$$

记住，由于 $\boldsymbol{X}$ 是 $n\times k$ 矩阵，而 $\boldsymbol{\beta}$ 是 $n\times 1$ 向量，所以 $\boldsymbol{X\beta}$ 是 $n\times 1$ 向量。

像在 3.2 节中一样，对 $\boldsymbol{\beta}$ 的估计要从最小化残差平方和开始。对任意可能的 $k\times 1$ 参数向量 $\boldsymbol{b}$，定义残差平方和为

$$\mathrm{SSR}(\boldsymbol{b}) = \sum_{t=1}^{n}(y_t - \boldsymbol{x}_t\boldsymbol{b})^2$$

普通最小二乘估计值的 $k\times 1$ 向量 $\hat{\boldsymbol{\beta}} = (\hat\beta_1,\hat\beta_2,\cdots,\hat\beta_k)'$ 对所有可能的 $k\times 1$ 向量 $\boldsymbol{b}$ 最小化了 SSR（$\boldsymbol{b}$）。这是多元微积分中的一个问题。最小化残差平方和的 $\hat{\boldsymbol{\beta}}$，一定是**一阶条件**（first order condition）

$$\partial\mathrm{SSR}(\hat{\boldsymbol{\beta}})/\partial\boldsymbol{b} \equiv 0 \tag{E.4}$$

的解。利用 $(y_t - \boldsymbol{x}_t\boldsymbol{b})^2$ 对 $\boldsymbol{b}$ 的导数是 $1\times k$ 向量 $-2(y_t - \boldsymbol{x}_t\boldsymbol{b})\boldsymbol{x}_t$ 的事实，式（E.4）等价于

$$\sum_{t=1}^{n}\boldsymbol{x}'_t(y_t - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}) \equiv \boldsymbol{0} \tag{E.5}$$

（已经除以 2 并取了转置。）我们可以把一阶条件写成

$$\sum_{t=1}^{n}(y_t - \hat\beta_1 - \hat\beta_2 x_{t2} - \cdots - \hat\beta_k x_{tk}) = 0$$

$$\sum_{t=1}^{n}x_{t2}(y_t - \hat\beta_1 - \hat\beta_2 x_{t2} - \cdots - \hat\beta_k x_{tk}) = 0$$

$$\cdots\cdots\cdots\cdots$$

$$\sum_{t=1}^{n}x_{tk}(y_t - \hat\beta_1 - \hat\beta_2 x_{t2} - \cdots - \hat\beta_k x_{tk}) = 0$$

除了标记习惯不同外，这与方程（3.13）中的一阶条件是一样的。我们想以

矩阵形式写出这些一阶条件，使之更有用。利用附录 D 中的分块矩阵公式，我们看出式（E.5）等价于

757

$$\mathbf{X}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{0} \tag{E.6}$$

或

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y} \tag{E.7}$$

可以证明，式（E.7）至少有一个解。多个解对我们并没有什么帮助，因为我们在求给定数据下惟一的 OLS 估计值集。假定 $k\times k$ 对称矩阵 $\mathbf{X}'\mathbf{X}$ 是非奇异的，那我们就可以将式（E.7）的两边同时左乘 $(\mathbf{X}'\mathbf{X})^{-1}$，以得到 OLS 估计量 $\hat{\boldsymbol{\beta}}$：

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{E.8}$$

这是对多元线性回归模型进行矩阵分析的关键公式。假定 $\mathbf{X}'\mathbf{X}$ 可逆就等于假定 rank（$\mathbf{X}$）$=k$，即 $\mathbf{X}$ 的列都是线性独立的。这也就是第 3 章中 MLR.4 的矩阵形式。

在我们继续讨论之前，要对式（E.8）给出一句警告。人们禁不住要对 $\hat{\boldsymbol{\beta}}$ 的这个公式简化如下：

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}=\mathbf{X}^{-1}(\mathbf{X}')^{-1}\mathbf{X}'\mathbf{y}=\mathbf{X}^{-1}\mathbf{y}$$

这个逻辑的疏漏之处在于，$\mathbf{X}$ 通常都不是方阵，所以就不可求逆。换句话说，除非 $n=k$（实践中这种情况几乎不可能发生），否则我们就不能写 $(\mathbf{X}'\mathbf{X})^{-1}=\mathbf{X}^{-1}(\mathbf{X}')^{-1}$。

OLS 拟合值和残差的 $n\times 1$ 向量由下式给出

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}},\ \hat{\mathbf{u}}=\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}$$

从式（E.6）和对 $\hat{\mathbf{u}}$ 的定义我们可以看到，$\hat{\boldsymbol{\beta}}$ 的一阶条件就是

$$\mathbf{X}'\hat{\mathbf{u}}=\mathbf{0} \tag{E.9}$$

由于 $\mathbf{X}$ 的第一列完全由 1 构成，所以式（E.9）意味着，当方程中包含了截距项时，残差和总为零，而且每个自变量和 OLS 残差之间的样本协方差也是零。（我们在第 3 章讨论了这两个性质。）

$$\mathrm{SSR}=\sum_{i=1}^{n}\hat{u}_i^2=\hat{\mathbf{u}}'\hat{\mathbf{u}}=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \tag{E.10}$$

第 3 章中所有的代数性质都可以用矩阵形式推导出来。比如，我们可以证明，总平方和等于解释平方和加上残差平方和［参见式（3.27）］。使用矩阵的证明并不比使用求和符号简单，所以我们就不再另外给出证明。

758 多元回归的矩阵方法可用做对回归作几何解释的基础。这里涉及的概念远比我们在附录 D 中所讨论的概念高深。［参见 Goldberger（1991）或 Greene（1997）。］

E.2 OLS 的有限样本性质

尽管矩阵代数能推导 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 的期望值和方差，但我们从假定开始就要小心从事。

假定 E.1（参数的线性）

模型可写成式（E.3）那样，其中 $\boldsymbol{y}$ 是一个可观测的 $n\times1$ 向量，$\boldsymbol{X}$ 是一个可观测的 $n\times k$ 矩阵，而 $\boldsymbol{u}$ 是不可观测的 $n\times1$ 误差或扰动项向量。

假定 E.2（零条件均值）

以整个矩阵 $\boldsymbol{X}$ 为条件，每个 u_t 的均值都为零：$\mathrm{E}(u_t\mid\boldsymbol{X})=0\ (t=1,2,\cdots,n)$。以向量形式即

$$\mathrm{E}(\boldsymbol{u}\mid\boldsymbol{X})=\boldsymbol{0} \tag{E.11}$$

在随机抽样假定 MLR.2 下，MLR.3 蕴涵了这个假定。在时间序列应用中，假定 E.2 对解释变量施加了严格的外生性假定，我们在第 10 章曾详细讨论过。这就排除了未来值与 u_t 相关的解释变量；具体而言，它淘汰了滞后因变量。在假定 E.2 下，我们在计算 $\hat{\boldsymbol{\beta}}$ 的期望值时可以 x_{tj} 为条件。

假定 E.3（不存在完全共线性）

矩阵 $\boldsymbol{X}$ 的秩为 k。

这是对排除解释变量之间线性相关假定的小心表述。在假定 E.3 下，$\boldsymbol{X}'\boldsymbol{X}$ 是非奇异的，所以 $\hat{\boldsymbol{\beta}}$ 是惟一的并可写为如式（E.8）。

定理 E.1（OLS 的无偏性）

在假定 E.1，E.2 和 E.3 下，OLS 估计量 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的无偏估计。

证明：利用假定 E.1 和 E.3 并经过简单运算，得到

759

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u})\\ &=(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\\ &=\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\end{aligned} \tag{E.12}$$

其中我们用到 $(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})=\boldsymbol{I}_k$。以 $\boldsymbol{X}$ 为条件取期望，得到

$$\mathrm{E}(\hat{\boldsymbol{\beta}}\mid\boldsymbol{X})=\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{u}\mid\boldsymbol{X})=\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{0}=\boldsymbol{\beta}$$

因为在假定 E.2 下，$\mathrm{E}(\boldsymbol{u}\mid\boldsymbol{X})=\boldsymbol{0}$。这个证明显然与 $\boldsymbol{\beta}$ 无关，所以我们已经证明 $\hat{\boldsymbol{\beta}}$ 是无偏的。

为了得到 $\hat{\boldsymbol{\beta}}$ 最简单形式的方差—协方差矩阵，我们施加同方差性和不存在序列相关的假定。

假定 E.4(同方差性和不存在序列相关)

(i) $\mathrm{Var}(u_t|\boldsymbol{X})=\sigma^2$, $t=1, 2, \cdots, n$。

(ii) $\mathrm{Cov}(u_t, u_s|\boldsymbol{X})=0$ 对任意的 $t\neq s$ 都成立。以矩阵形式，我们可以将这两个假定写成

$$\mathrm{Var}(\boldsymbol{u}|\boldsymbol{X})=\sigma^2\boldsymbol{I}_n \tag{E.13}$$

式中，$\boldsymbol{I}_n$ 为 $n\times n$ 单位矩阵。

假定 E.4 中的第 (i) 部分是同方差性假定：u_t 的方差不能依赖于 $\boldsymbol{X}$ 中的任何一个元素，而且不同观测 t 的方差都相等。第 (ii) 部分是不存在序列相关的假定：不同观测的误差不能相关。在随机抽样的情况下，以及在其他独立观测的横截面抽样方案中，假定 E.4 的第 (ii) 部分自动成立。而对于时间序列的应用研究而言，第 (ii) 部分排除了不同时期的误差之间存在相关的可能性（无论是以 $\boldsymbol{X}$ 为条件还是无条件）。

由于式 (E.13)，我们通常在假定 E.4 成立时称 $\boldsymbol{u}$ 具有**数量方差—协方差矩阵** (scalar variance-covariance matrix)。我们现在可以推导 OLS 估计量的方差—协方差矩阵。

定理 E.2(OLS 估计量的方差—协方差矩阵)

在假定 E.1~E.4 下，

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{E.14}$$

证明：从方程 (E.12) 中的最后一个公式，我们有

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})&=\mathrm{Var}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}|\boldsymbol{X}]\\&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\mathrm{Var}(\boldsymbol{u}|\boldsymbol{X})]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

现在我们利用假定 E.4 得到

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I}_n)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\&=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1}=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

760 公式 (E.14) 意味着，$\hat{\beta}_j$ 的方差（以 $\boldsymbol{X}$ 为条件）由 σ^2 乘以 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的主对角线上第 j 个元素而得到。至于斜率系数，我们在方程 (3.51) 中给出了一个可解释的公式。方程 (E.14) 还告诉我们如何得到任意两个 OLS 估计值之间的协方差：将 σ^2 与 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 主对角线以外的某个适当元素相乘。我们在第 4 章证明了，如何通过适当重写模型，而无须明确求出得到置信区间和假设检验所需要的协方差。

现在可以证明很一般性的高斯-马尔科夫定理了。

定理 E.3（高斯-马尔科夫定理）

在假定 E.1～E.4 下，$\boldsymbol{\beta}$ 是最优线性无偏估计量。

证明：$\boldsymbol{\beta}$ 的其他任何一个线性估计量都可以写成

$$\boldsymbol{\beta} = \boldsymbol{A}'\boldsymbol{y} \tag{E.15}$$

式中，$\boldsymbol{A}$ 为一个 $n \times k$ 矩阵。为了使 $\tilde{\boldsymbol{\beta}}$ 成为以 $\boldsymbol{X}$ 为条件的无偏估计量，$\boldsymbol{A}$ 必须由非随机数字和 $\boldsymbol{X}$ 的函数构成。（例如 $\boldsymbol{A}$ 就不能是 $\boldsymbol{y}$ 的函数。）为了看出所需要的对 $\boldsymbol{A}$ 的进一步限制，写出

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{A}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) = (\boldsymbol{A}'\boldsymbol{X})\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{u} \tag{E.16}$$

于是

$$\begin{aligned} \mathrm{E}(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{A}'\boldsymbol{u} \mid \boldsymbol{X}) \\ &= \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}'\mathrm{E}(\boldsymbol{u} \mid \boldsymbol{X}) && \text{因为 } \boldsymbol{A} \text{ 是 } \boldsymbol{X} \text{ 的一个函数} \\ &= \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} && \text{因为 } \mathrm{E}(\boldsymbol{u} \mid \boldsymbol{X}) = \boldsymbol{0} \end{aligned}$$

为了使 $\tilde{\boldsymbol{\beta}}$ 成为 $\boldsymbol{\beta}$ 的无偏估计量，必须有 E（$\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}$）$= \boldsymbol{\beta}$ 对所有的 $k \times 1$ 向量都成立，即

$$\boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta} \quad \text{对所有 } k \times 1 \text{ 向量 } \boldsymbol{\beta} \tag{E.17}$$

由于 $\boldsymbol{A}'\boldsymbol{X}$ 是一个 $k \times k$ 矩阵，所以当且仅当 $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$ 时式（E.17）成立。方程（E.15）和方程（E.17）就刻画了 $\boldsymbol{\beta}$ 的所有线性无偏估计量的特征。

接下来，我们根据假定 E.4，从式（E.16）得到

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \boldsymbol{A}'[\mathrm{Var}(\boldsymbol{u} \mid \boldsymbol{X})]\boldsymbol{A} = \sigma^2\boldsymbol{A}'\boldsymbol{A}$$

$$\begin{aligned} \text{因此，} \mathrm{Var}(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) - \mathrm{Var}(\boldsymbol{\beta} \mid \boldsymbol{X}) &= \sigma^2[\boldsymbol{A}'\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1}] \\ &= \sigma^2[\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{A}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{A}] && \text{因为 } \boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k \\ &= \sigma^2\boldsymbol{A}'[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{A} \\ &= \sigma^2\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A} \end{aligned}$$

式中，$\boldsymbol{M} \equiv \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$。由于 $\boldsymbol{M}$ 是对称等幂矩阵，所以 $\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A}$ 对任意 $n \times k$ 矩阵 $\boldsymbol{A}$ 都是半正定的。这就证明了，OLS 估计量 $\boldsymbol{\beta}$ 是 BLUE。这一点有什么重要意义呢？令 $\boldsymbol{c}$ 表示任意一个 $k \times 1$ 向量，并考虑一个线性组合 $\boldsymbol{c}'\boldsymbol{\beta} = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k$，它是一个数量。$\boldsymbol{c}'\boldsymbol{\beta}$ 的无偏估计量是 $\boldsymbol{c}'\boldsymbol{\beta}$ 和 $\boldsymbol{c}'\tilde{\boldsymbol{\beta}}$。但由于［Var（$\boldsymbol{c}\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}$）－Var（$\boldsymbol{c}\boldsymbol{\beta} \mid \boldsymbol{X}$）］是半正定的，所以

$$\mathrm{Var}(\boldsymbol{c}\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) - \mathrm{Var}(\boldsymbol{c}\boldsymbol{\beta} \mid \boldsymbol{X}) = \boldsymbol{c}'[\mathrm{Var}(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) - \mathrm{Var}(\boldsymbol{\beta} \mid \boldsymbol{X})]\boldsymbol{c} \geqslant 0$$

因此当用来估计 $\boldsymbol{\beta}$ 的任何一个线性组合时，OLS 总得到最小方差。具体而言，对 β_j 的其他任何线性无偏估计量，都有 Var（$\boldsymbol{\beta}_j \mid \boldsymbol{X}$）$\leqslant$ Var（$\tilde{\boldsymbol{\beta}}_j \mid \boldsymbol{X}$）。

误差方差 σ^2 的无偏估计量可写成

$$\hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)$$

因为我们的记号已经标明，包括截距项在内总共 k 个参数。

定理 E.4（$\hat{\sigma}^2$ 的无偏性）

在假定 E.1～E.4 下，$\hat{\sigma}^2$ 是 σ^2 的无偏估计：对所有的 $\sigma^2>0$，都有 $\mathrm{E}(\hat{\sigma}^2|\mathbf{X})=\sigma^2$。

证明：写下 $\hat{\mathbf{u}}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}=\mathbf{M}\mathbf{y}=\mathbf{M}\mathbf{u}$，其中 $\mathbf{M}=\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$，且最后一个等式成立是因为 $\mathbf{M}\mathbf{X}=\mathbf{0}$。因为 $\mathbf{M}$ 是对称等幂矩阵，所以

$$\hat{\mathbf{u}}'\hat{\mathbf{u}}=\mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u}=\mathbf{u}'\mathbf{M}\mathbf{u}$$

因为 $\mathbf{u}'\mathbf{M}\mathbf{u}$ 是一个数量，所以它就等于它的迹。因此

$$\begin{aligned}\mathrm{E}(\mathbf{u}'\mathbf{M}\mathbf{u}|\mathbf{X}) &= \mathrm{E}[\mathrm{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})|\mathbf{X}]=\mathrm{E}[\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}')|\mathbf{X}]\\ &=\mathrm{tr}[\mathrm{E}(\mathbf{M}\mathbf{u}\mathbf{u}')|\mathbf{X}]=\mathrm{tr}[\mathbf{M}\mathrm{E}(\mathbf{u}\mathbf{u}')|\mathbf{X}]\\ &=\mathrm{tr}(\mathbf{M}\sigma^2\mathbf{I}_n)=\sigma^2\mathrm{tr}(\mathbf{M})\\ &=\sigma^2(n-k)\end{aligned}$$

最后一个等式成立是因为 $\mathrm{tr}(\mathbf{M})=\mathrm{tr}(\mathbf{I}_n)-\mathrm{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')=n-\mathrm{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})=n-\mathrm{tr}(\mathbf{I}_k)=n-k$。因此

$$\mathrm{E}(\hat{\sigma}^2|\mathbf{X})=\mathrm{E}(\mathbf{u}'\mathbf{M}\mathbf{u}|\mathbf{X})/(n-k)=\sigma^2$$

E.3 统计推断

当我们增加最后一个经典线性模型假定后，$\hat{\boldsymbol{\beta}}$ 就具有多元正态分布，这就导致第 4 章讨论的标准检验统计量所要求的 t 和 F 分布。

假定 E.5（误差的正态性）

以 $\mathbf{X}$ 为条件，u_t 服从独立同分布 Normal$(0,\sigma^2)$。换言之，给定 $\mathbf{X}$ 下的 $\mathbf{u}$ 服从均值为 0 和方差—协方差矩阵为 $\sigma^2\mathbf{I}_n$ 的多元正态分布：$\mathbf{u}\sim\mathrm{Normal}(\mathbf{0},\sigma^2\mathbf{I}_n)$。

762 在假定 E.5 下，每个 u_t 都独立于解释变量。在时间序列的背景中，这实质上是一个严格外生性假定。

定理 E.5（$\hat{\boldsymbol{\beta}}$ 的正态性）

在经典线性模型假定 E.1～E.5 下，以 $\mathbf{X}$ 为条件的 $\hat{\boldsymbol{\beta}}$ 服从均值为 $\boldsymbol{\beta}$ 和方差—协方差矩阵为 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 的多元正态分布。

定理 E.5 是进行涉及 $\hat{\boldsymbol{\beta}}$ 的统计推断的基础。事实上，与我们在附录 D 中

概述的 χ 平方、t 和 F 分布的性质一起，我们可以用定理 E.5 证明，t 统计量在假定 E.1～E.5 下（在虚拟假设下）具有 t 分布，F 统计量的推导证明也类似。我们用对 t 统计量的证明来加以说明。

定理 E.6

在假定 E.1～E.5 下，有

$$(\hat{\beta}_j-\beta_j)/\mathrm{se}(\hat{\beta}_j)\sim t_{n-k},\quad j=1,2,\cdots,k$$

证明：证明过程要分几步进行，如下证明都以 $\boldsymbol{X}$ 为条件。首先，根据定理 E.5，有 $(\hat{\beta}_j-\beta_j)/\mathrm{sd}(\hat{\beta}_j)\sim\mathrm{Normal}(0,1)$，其中 $\mathrm{sd}(\hat{\beta}_j)=\sigma\sqrt{c_{jj}}$，而 c_{jj} 是 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 中主对角线上第 j 个元素。然后在假定 E.1～E.5 下，以 $\boldsymbol{X}$ 为条件，有

$$(n-k)\,\hat{\sigma}^2/\sigma^2\sim\chi^2_{n-k}\tag{E.18}$$

这是因为 $(n-k)\,\hat{\sigma}^2/\sigma^2=(\boldsymbol{u}/\sigma)'\boldsymbol{M}(\boldsymbol{u}/\sigma)$，其中 $\boldsymbol{M}$ 是定理 E.4 中定义的 $n\times n$ 对称等幂矩阵。但根据假定 E.5，$\boldsymbol{u}/\sigma\sim\mathrm{Normal}(\boldsymbol{0},\boldsymbol{I}_n)$。于是根据附录 D 中 χ 平方分布的性质 1，就得到 $(\boldsymbol{u}/\sigma)'\boldsymbol{M}(\boldsymbol{u}/\sigma)\sim\chi^2_{n-k}$（因为 $\boldsymbol{M}$ 的秩为 $n-k$）。

我们还需要证明 $\boldsymbol{\beta}$ 和 $\hat{\sigma}^2$ 是独立的。但 $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$，而 $\hat{\sigma}^2=\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}/(n-k)$。现在，由于 $\boldsymbol{X}'\boldsymbol{M}=\boldsymbol{0}$，所以 $[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{M}=\boldsymbol{0}$。于是根据附录 D 中多元正态分布的性质 5，得知 $\boldsymbol{\beta}$ 和 $\boldsymbol{M}\boldsymbol{u}$ 是独立的。由于 $\hat{\sigma}^2$ 是 $\boldsymbol{M}\boldsymbol{u}$ 的一个函数，所以 $\boldsymbol{\beta}$ 和 $\hat{\sigma}^2$ 也是独立的。

最后，我们可以写出

$$(\hat{\beta}_j-\beta_j)/\mathrm{se}(\hat{\beta}_j)=[(\hat{\beta}_j-\beta_j)/\mathrm{sd}(\hat{\beta}_j)]/(\hat{\sigma}^2/\sigma^2)^{1/2}$$

它是一个标准正态随机变量和一个 $\chi^2_{n-k}/(n-k)$ 随机变量的平方根之比。我们刚刚证明了它们是独立的，所以根据对 t 随机变量的定义，$(\hat{\beta}_j-\beta_j)/\mathrm{se}(\hat{\beta}_j)$ 服从 t_{n-k} 分布。由于这个分布与 $\boldsymbol{X}$ 无关，所以它也是 $(\hat{\beta}_j-\beta_j)/\mathrm{se}(\hat{\beta}_j)$ 的无条件分布。

通过这个定理，我们可以代入任意一个 β_j 的假设值，并像通常那样利用 t 统计量进行假设检验。

在假定 E.1～E.5 下，我们可以针对 $\boldsymbol{\beta}$ 的无偏估计量（同样以 $\boldsymbol{X}$ 为条件）的方差—协方差矩阵计算所谓的 Cramer-Rao 下界［参见 Greene（1997，第 4 章）］。可以证明它等于 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$，恰好就是 OLS 估计量的方差—协方差矩阵。这就意味着，$\boldsymbol{\beta}$ 是 $\boldsymbol{\beta}$ 的**最小方差无偏**（minimum variance unbiased）估计量（以 $\boldsymbol{X}$ 为条件）：对其他任何一个无偏估计量 $\tilde{\boldsymbol{\beta}}$，$\mathrm{Var}(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})-\mathrm{Var}(\boldsymbol{\beta}|\boldsymbol{X})$ 都是半正定矩阵；我们无须只限于考虑对 $\boldsymbol{y}$ 为线性的估计量。

很容易证明，在假定 E.5 下，OLS 估计量实际上是 $\boldsymbol{\beta}$ 的最大似然估计量。对每个 t，给定 $\boldsymbol{X}$ 下 y_t 的分布是 Normal（$\boldsymbol{x}_t\boldsymbol{\beta}$，$\sigma^2$）。由于 y_t 独立地以

$\boldsymbol{X}$ 为条件，所以样本的似然函数可从密度之积得到：

$$\prod_{t=1}^{n}(2\pi\sigma^2)^{1/2}\exp[-(y_t-\boldsymbol{x}_t\boldsymbol{\beta})^2/(2\sigma^2)]$$

将这个函数对 $\boldsymbol{\beta}$ 和 σ^2 最大化，就等于最大化其自然对数：

$$\sum_{t=1}^{n}[-(1/2)\log(2\pi\sigma^2)-(y_t-\boldsymbol{x}_t\boldsymbol{\beta})^2/(2\sigma^2)]$$

为了得到 $\boldsymbol{\beta}$，这就等于在最小化 $\sum_{t=1}^{n}(y_t-\boldsymbol{x}_t\boldsymbol{\beta})^2$（除以 $2\sigma^2$ 不影响最优化），这正是 OLS 所求解的问题。我们所用的 σ^2 的估计量 SSR/（$n-k$），实际上不是 σ^2 的 MLE；这个 MLE 是一个有偏误的估计量 SSR/n。由于 σ^2 的无偏估计量导致 t 和 F 统计量在虚拟假设下恰好具有 t 和 F 分布，所以我们总是使用这个无偏估计量而不是用 MLE。

▶ 小结

本附录简要讨论了用矩阵方法表述的线性回归模型。虽然为使用矩阵代数的较高年级准备了这些内容，但不需要先读它再读正文中的内容。事实上，本附录证明的一些结论，我们曾在正文中或者未经证明而直接表述，或者只在特例情形下证明，或者以前的证明方法更麻烦。使用矩阵代数后，可以对其他专题（如渐近性质、工具变量估计和综列数据模型等）进行更简明的处理。详细情况，可查阅包括戴维森和麦金农（Davidson and MacKinnon，1993）、格林（Greene，1997）和伍德里奇（Wooldridge，1999）在内的高级计量经济学教材。

关键术语

一阶条件
矩阵符号
最小方差无偏
数量方差—协方差矩阵
OLS 估计量的方差—协方差矩阵

习题

764 **E.1** 令 $\boldsymbol{x}_t$ 表示观测 t 的 $1\times k$ 解释变量向量。证明 OLS 估计量 $\boldsymbol{\beta}$ 可以

写成

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{x}_t' y_t\right)$$

通过将每个和都除以 n 证明 $\hat{\boldsymbol{\beta}}$ 是样本均值的一个函数。

E.2 令 $\hat{\boldsymbol{\beta}}$ 表示 OLS 估计值的 $k \times 1$ 向量。

(i) 证明对任何一个 $k \times 1$ 向量 $\boldsymbol{b}$，我们都可以将残差平方和写成

$$\mathrm{SSR}(\boldsymbol{b}) = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} + (\hat{\boldsymbol{\beta}} - \boldsymbol{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{B})$$

[提示：写出 $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = [\hat{\boldsymbol{u}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{b})]'[\hat{\boldsymbol{u}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{b})]$ 并利用 $\boldsymbol{X}'\hat{\boldsymbol{u}} = \boldsymbol{0}$ 的事实。]

(ii) 请解释：在假定 $\boldsymbol{X}$ 的秩为 k 后，第（i）部分中 SSR（$\boldsymbol{b}$）的表达式是如何证明了，$\hat{\boldsymbol{\beta}}$ 是所有可能的 $\boldsymbol{b}$ 值中惟一最小化 SSR（$\boldsymbol{b}$）的向量。

E.3 令 $\hat{\boldsymbol{\beta}}$ 表示 $\boldsymbol{y}$ 对 $\boldsymbol{X}$ 回归得到的 OLS 估计值。令 $\boldsymbol{A}$ 表示一个 $k \times k$ 非奇异矩阵并定义 $\boldsymbol{z}_t \equiv \boldsymbol{x}_t\boldsymbol{A}$（$t = 1, 2, \cdots, n$）。因此 $\boldsymbol{z}_t$ 是 $1 \times k$ 向量并且是一个非退化的 $\boldsymbol{x}_t$ 的线性组合。令 $\boldsymbol{Z}$ 表示以 $\boldsymbol{z}_t$ 为行的 $n \times k$ 矩阵。令 $\tilde{\boldsymbol{\beta}}$ 表示 $\boldsymbol{y}$ 对 $\boldsymbol{Z}$ 回归所得到的 OLS 估计值。

(i) 证明：$\tilde{\boldsymbol{\beta}} = \boldsymbol{A}^{-1}\hat{\boldsymbol{\beta}}$。

(ii) 令 $\hat{y}_t$ 表示原回归的拟合值，而令 $\tilde{y}_t$ 表示 $\boldsymbol{y}$ 对 $\boldsymbol{Z}$ 回归的拟合值。证明对所有 $t = 1, 2, \cdots, n$，都有 $\hat{y}_t = \tilde{y}_t$。这两个回归的残差相比如何？

(iii) 证明估计的 $\tilde{\boldsymbol{\beta}}$ 的方差矩阵为 $\hat{\sigma}^2\boldsymbol{A}^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{A}^{-1})'$，其中 $\hat{\sigma}^2$ 是从 $\boldsymbol{y}$ 对 $\boldsymbol{X}$ 的回归中所得到的通常的方差估计值。

(iv) 令 $\hat{\beta}_j$ 表示 y_t 对 $1, x_{t2}, \cdots, x_{tk}$ 回归所得到的 OLS 估计值，令 $\tilde{\beta}_j$ 表示 y_t 对 $1, a_2x_{t2}, \cdots, a_kx_{tk}$ 回归所得到的 OLS 估计值，其中 $a_j \neq 0$（$j = 2, \cdots, k$）。利用第（i）部分的结论求出 $\hat{\beta}_j$ 和 $\tilde{\beta}_j$ 之间的关系。

(v) 假定第（iv）部分中的假设，使用第（iii）部分证明 se（$\tilde{\beta}_j$）= se（$\hat{\beta}_j$）/$|a_j|$。

(vi) 假定第（iv）部分中的假设，证明 $\hat{\beta}_j$ 和 $\tilde{\beta}_j$ 的 t 统计量的绝对值相等。

附录 F　各章习题解答

第2章

765

习题 2.1

当学生能力、学习动机、年龄和 u 中的其他因素与到课率无关时，式(2.6)将成立。这看起来不太像那回事。

习题 2.2

约9.64美元。为了看出这一点，从1976年以1997年美元度量的平均工资，我们可以得到CPI缩减指数为16.64/5.90≈2.82。将3.42乘以2.82就得到9.64。

习题 2.3

将 $shareA=60$ 代入方程(2.28)就可以看到它等于59.26。这并非不合理：如果候选人A花了总竞选支出的60%，那么预测他或她将得到刚好高于

59%的选票。

习题 2.4

通过将方程(2.39)乘以 10 很容易看出，这个方程是 $\widehat{salaryhun} = 9\,631.91 + 185.01\,roe$。

习题 2.5

方程(2.28)可写成 $\mathrm{Var}(\hat{\beta}_0) = (\sigma^2 n^{-1})(\sum_{i=1}^{n} x_i^2)/(\sum_{i=1}^{n}(x_i - \bar{x})^2)$，其中与 $\sigma^2 n^{-1}$ 相乘的项大于或等于 1，但当且仅当 $\bar{x} = 0$ 时等于 1。此时，方差达到它可能达到的最小值：$\mathrm{Var}(\hat{\beta}_0) = \sigma^2/n$。

第 3 章

习题 3.1

只有少数因素包括了年龄和性别分布、警力规模(或更一般地，投入到与犯罪做斗争的资源)、人口和一般历史因素。这些因素当然可能与 *prbconv* 和 766 *avgsen* 相关，这就意味着式(3.5)不成立。比如，某些在预防犯罪和执法方面投入较多气力的城市，其警力规模可能与 *prbconv* 和 *avgsen* 都相关。

习题 3.2

我们利用 OLS 与预测值和残差有关的第三个性质：当在 OLS 回归线中代入所有自变量的平均值时，就得到因变量的平均值。所以 $\overline{colGAP} = 1.29 + 0.453\,\overline{hsGPA} + 0.009\,4\,\overline{ACT} = 1.29 + 0.453(3.4) + 0.009\,4(24.2) \approx 3.06$。你可以在 GPA1.RAW 中验证 *colGPA* 的平均值，从而在保留两位小数的情况下证实这一点。

习题 3.3

否。变量 *shareA* 不是 *expendA* 和 *expendB* 的一个完全线性函数，相反它是一个非线性的函数：$shareA = 100 \cdot [expendA/(expendA + expendB)]$。因此，将 *expendA*，*expendB* 和 *shareA* 作为解释变量是合理的。

习题 3.4

如我们在 3.4 节所讨论的那样，如果我们关心 x_1 对 y_1 的影响，那么，其他解释变量之间(x_2，x_3 等)的相关关系并不影响 $\mathrm{Var}(\hat{\beta}_1)$。这些变量作为控制变量而包括进来，我们不必担心这种共线性。当然，我们控制它们的主要原因是，虽然我们认为它们与到课率相关，但这是进行其他条件不变的分析所必要的。

第 4 章

习题 4.1

在这些假定下，高斯-马尔科夫假定是满足的：由于 u 独立于解释变量，所以 $\mathrm{E}(u|x_1,\cdots,x_k)=\mathrm{E}(u)$ 和 $\mathrm{Var}(u|x_1,\cdots,x_k)=\mathrm{Var}(u)$。而且，很容易看出$\mathrm{E}(u)=0$。因此，MLR.3 和 MLR.5 成立。因为 u 不是正态分布的（违背了 MLR.6），所以不满足经典线性模型假定。

习题 4.2

$\mathrm{H}_0:\beta_1=0$，$\mathrm{H}_1:\beta_1<0$。

习题 4.3

因为 $\hat{\beta}_1=0.56>0$，而我们正在相对 $\mathrm{H}_1:\beta_1>0$ 进行检验，所以单侧 p 值是双侧 p 值的一半，或 0.043。

习题 4.4

$\mathrm{H}_0:\beta_5=\beta_6=\beta_7=\beta_8=0$，$k=8$ 和 $q=4$。受约束形式的模型为

$$score=\beta_0+\beta_1 classize+\beta_2 expend+\beta_3 tchcomp+\beta_4 enroll+u$$

767 ### 习题 4.5

检验排除 ACT 的 F 统计量是 $[(0.291-0.183)/(1-0.291)](680-3)\approx 103.13$。因此，$t$ 统计量的绝对值约为 10.16。因为 $\hat{\beta}_{ACT}$ 为负，所以 ACT 的 t 统计量为负，于是 $t_{ACT}=-10.16$。

习题 4.6

影响不大。*droprate* 和 *gradrate* 联合显著性的 F 检验很容易从表中的 R-平方计算出来：$F=[(0.361-0.353)/(1-0.361)](402/2)\approx 2.52$。从表 G.3（a）得到 10% 的临界值为 2.30，而从表 G.3（b）得到 5% 的临界值为 3。p 值约为 0.082。因此，*droprate* 和 *gradrate* 在 10% 的显著性水平上是联合显著的，但在 5% 的水平上不是。无论如何，控制这些因素对 b/s 的系数都只具有次要的影响。

第 5 章

习题 5.1

这就需要某些假定。看起来假定 $\beta_2>0$（*priGPA* 对 *score* 有正影响）和 Cov（*skipped*，*priGPA*）<0（*skipped* 和 *priGPA* 负相关）是合理的。这意味着 $\beta_2\delta_1<0$，从而意味着 $\tilde{\beta}_1<\beta_1$。由于认为 β_1 为负，所以一个简单回归很可能高估缺课的重要性。

习题 5.2

$\hat{\beta}_j \pm 1.96\mathrm{se}(\hat{\beta}_j)$ 是渐近的 95% 置信区间。或者，可以把 1.96 换成 2。

第 6 章

习题 6.1

由于 $fincdol=1\,000\cdot faminc$，所以 *fincdol* 的系数是 *faminc* 的系数除以 1 000，或0.092 7/1 000＝0.000 092 7。标准误也下降1 000倍，所以 t 统计量不变，其他任何一个 OLS 统计量也都不变。为便于阅读，最好以千美元度量家庭收入。

习题 6.2

我们可以做一般性的分析。方程为

$$\log(y)=\beta_0+\beta_1\log(x_1)+\beta_2x_2+\cdots$$

式中，x_2 表示一个比例而不是百分数。于是，在其他条件不变的情况下，$\Delta\log(y)=\beta_2\Delta x_2$，$100\cdot\Delta\log(y)=\beta_2(100\cdot\Delta x_2)$ 或 $\%\Delta y\approx\beta_2(100\cdot\Delta x_2)$。现在，因为 Δx_2 是比例的变化，所以 $100\cdot\Delta x_2$ 就是百分数的变化。特别是，若 $\Delta x_2=0.01$，则 $100\cdot\Delta x_2=1$，这就对应于一个百分点变化。但当 $100\cdot\Delta x_2=1$ 时，β_2 就是百分比变化。

768
习题 6.3

新模型是 $stndfnl=\beta_0+\beta_1atndrte+\beta_2priGPA+\beta_3ACT+\beta_4priGPA^2+\beta_5ACT^2+\beta_6priGPA\cdot atndrte+\beta_7ACT\cdot atndrte+u$。

因此，*atndrte* 对 *stndfnl* 的偏效应是 $\beta_1+\beta_6priGPA+\beta_7ACT$。这就是我们为得到 *stndfnl* 在其他条件不变情况下的变化，而乘以 $\Delta atndrte$ 的倍数。

习题 6.4

从方程（6.21）有 $\overline{R}^2 = 1 - \hat{\sigma}^2 / [SST/(n-1)]$。对于一个给定样本和一个给定的因变量，$SST/(n-1)$ 是固定的。当我们使用不同的解释变量集时，只有 $\hat{\sigma}^2$ 发生变化。$\overline{R}^2$ 随着 $\hat{\sigma}^2$ 的下降而提高。如果使 $\hat{\sigma}$ 因而使 $\hat{\sigma}^2$ 尽可能小，那我们就会使 $\overline{R}^2$ 尽可能大。

习题 6.5

一种可能性是搜集一个演员样本的年收入及其所演电影获利能力方面的数据。在一个简单的回归分析中，我们可以将收入与获利能力相关。但我们可以应该控制可能影响薪水的其他因素，如年龄、性别及所演电影的种类。在回归模型中包括定性变量的方法将在第 7 章考虑。

第 7 章

习题 7.1

不是，因为它没有明确 *party* 什么时候取值 1 而什么时候取值 0。一个更好的名称是像 *Dem* 那样的变量，对民主党取值 1 而对共和党取值 0。或者是对共和党取值 1 而对民主党取值 0 的 *Rep*。

习题 7.2

以 *outfield* 为基组，我们将包括虚拟变量 *frstbase*，*part*，*scndbase*，*shrtstop* 和 *catcher*。

习题 7.3

这种情况下的虚拟假设是 H_0：$\delta_1 - \delta_2 - \delta_3 - \delta_4 - 0$，所以有四个约束。和通常一样，我们将使用一个 F 检验（其中，$q = 4$，而 k 取决于其他解释变量的个数）。

习题 7.4

由于 *tenure* 以二次形式出现，所以我们应该容许男性和女性分别有一个二次形式，即我们增加解释变量 *female*·*tenure* 和 *female*·*tenure*^{2}。

习题 7.5

769 在式（7.31）中代入 $pcnv = 0$，$avgsen = 0$，$tottime = 0$，$ptime86 = 0$，$qemp86 = 0$，$black = 1$ 和 $hispan - 0$：$\widehat{arr}86 = 0.380 - 0.038\ (4)\ + 0.170 = 0.398$ 或几乎 0.4。很难知道这是否“合理”。对于那些全年处在就业中而又没有定过罪的人来说，这个估计值可能看起来高了一些，但要记住，总体是

由那些在 1986 年以前至少被拘捕一次的人组成的。

第 8 章

习题 8.1

这个说法显然是错的。比如，在方程（8.7）中，*black* 通常的标准误是 0.147，而异方差稳健标准误是 0.118。

习题 8.2

F 检验将通过 $\hat{u}^2$ 对 *marrmale*、*marrfem* 和 *singfem*（*singmale* 是基组）的回归而得到。在 $n = 526$ 和回归中有三个自变量时，两个 df 是 3 和 522。

习题 8.3

实际上没有。因为这是一个简单的回归模型，所以异方差性只在它与 *inc* 相关时才成问题。但这种情形中的 Breusch-Pagan 检验等价于 $\hat{u}^2$ 对 *inc* 的回归中的 t 统计量。一个大小为 0.96 的 t 统计量不足以拒绝同方差性假定。

习题 8.4

我们除了计算对异方差稳健的标准误之外仍可以使用加权最小二乘法。在方程（8.26）中，如果我们的方差模型不正确，那我们仍将有异方差性。因此，我们可以猜测异方差性的形式并使用 WLS，但我们的分析可以对不正确的异方差性形式说是稳健的。不幸的是，我们可能必须明确得到变换后的变量。

第 9 章

习题 9.1

有些二值变量，将它们平方后没有影响：$black^2 = black$ 和 $hispan^2 = hispan$。

习题 9.2

当 $educ \cdot IQ$ 在方程中时，*educ* 的系数（比方说 β_1）度量了 *educ* 在 $IQ = 0$ 时对 log（*wage*）的影响。（教育的偏效应是 $\beta_1 + \beta_9 IQ$）所考虑的总体中没有一个人的 IQ 接近于零。在总体平均 IQ 为 100 时，从第（3）列估计的教育回报是 $0.018 + 0.000\,34(100) = 0.052$，约等于我们在第（2）列中

得到的系数。

习题 9.3

否。如果 $educ^*$ 是一个整数（意味着一个人在刚读完的年级后没有接受教育），则测量误差是零。如果 $educ^*$ 不是一个整数，则 $educ < educ^*$，所以测量误差为负。起码，e_1 的均值不可能为零，而且 e_1 和 $educ^*$ 可能相关。

770

习题 9.4

一个在任总统不参加竞选的决策可以与他或她预期在竞选中如何去做系统相关。因此，我们的样本可能只是一个比所有可能参加竞选的在任者更强的在任者样本。如果所关心的总体包括所有的在任者，那么这样就会导致样本选择问题。如果我们只关心追求连任者的竞选支出对选举结果的影响，那就不存在样本选择问题。

第 10 章

习题 10.1

即期倾向是 0.48，长期倾向是 $0.48 - 0.15 + 0.32 = 0.65$。

习题 10.2

解释变量是 $x_{t1} = z_t$，$x_{t2} = z_{t-1}$。没有完全共线性意味着这些变量不会是常数，而且样本中的这些变量不会有线性关系。这一点就排除了 $z_1, \cdots, z_n$ 取相同值或 $z_0, z_1, \cdots, z_{n-1}$ 取相同值的可能性。同时，无完全共线性也排除了其他形式，诸如，如果 $z_t = a + bt$（a 和 b 为常数），那么，$z_{t-1} = a + b(t-1) = (a + bt) - b = z_t - b$，它就是 z_t 的一个完全共线性函数。

习题 10.3

如果 $\{z_t\}$ 随着时间缓慢变化——很多经济时间序列的水平值或对数值就是如此——那么 z_t 和 z_{t-1} 就是高度相关的。比如 PHLLIPS.RAW 中的 $unem_t$ 和 $unem_{t-1}$ 的相关程度为 0.74。

习题 10.4

不会。因为随着时间 t 的增大，当 $\alpha_1 < 0$ 时线性时间趋势（负数）就会变得越来越负得多。但是，因为 gfr 不可能是负的，所以有负的趋势系数的时间趋势不可能正确表示出所有将来时期的 gfr。

习题 10.5

3 月份的截距是 $\beta_0 + \delta_2$。季节性虚拟变量是严格外生的，这是因为它们

遵循某种确定的模式。比如说，月份是不随着解释变量或因变量的变化而变化的。

第 11 章

习题 11.1

(i) 不是，因为 E（y_t）$= \delta_0 + \delta_1 t$ 与 t 的大小有关。

(ii) 是，因为 $y_t - \mathrm{E}$（y_t）$= e_t$ 是一个 $i.i.d.$ 序列。

771 ### 习题 11.2

把 $inf_t^e =$（1/2）$inf_{t-1} +$（1/2）inf_{t-2} 代入 $inf_t - inf_t^e = \beta_1$（$unem_t - \mu_0$）$+ e_t$，并整理得到 $inf_t -$（1/2）（$inf_{t-1} + inf_{t-2}$）$= \beta_0 + \beta_1 unem_t + e_t$，其中 $\beta_0 = -\beta_1 \mu_0$。于是，我们可以做 y_t 对 $unem_t$ 的回归，其中。注意 $y_t = inf_t -$（1/2）（$inf_{t-1} + inf_{t-2}$），如此构造 y_t 后我们就失去了头两次观测值。

习题 11.3

不是的。因为 u_t 和 u_{t-1} 是相关的。特别是，如果 $\alpha_1 \neq 0$，就有 Cov（u_t，u_{t-1}）$= \mathrm{E}$（$e_t + \alpha_1 e_{t-1}$）（$e_{t-1} + \alpha_1 e_{t-2}$）$= \alpha_1 \mathrm{E}$（$e_{t-1}^2$）$= \alpha_1 \sigma_e^2 \neq 0$。于是，倘若这里的误差是序列相关的，这个模型就不是动态完整的。

第 12 章

习题 12.1

利用方程（12.4）。在这里，只有相邻项是相关的。$x_t u_t$ 和 $x_{t+1} u_{t+1}$ 之间的协方差是 $x_t x_{t+1}$Cov（u_t，u_{t+1}）$= x_t x_{t+1} \alpha \sigma_e^2$。所以，公式为

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}_1) &= \mathrm{SST}_x^{-2}\Big[\sum_{t=1}^{n} x_t^2 \mathrm{Var}(u_t) + 2\sum_{t=1}^{n-1} x_t x_{t+1} \mathrm{E}(u_t u_{t+1})\Big] \\
&= \sigma^2/\mathrm{SST}_x + (2/\mathrm{SST}_x^2)\sum_{t=1}^{n-1} \alpha \sigma_e^2 x_t x_{t+1} \\
&= \sigma^2/\mathrm{SST}_x + \alpha \sigma_e^2 (2/\mathrm{SST}^2 x)\sum_{t=1}^{n-1} x_t x_{t+1}
\end{aligned}
$$

式中，$\sigma^2 = \mathrm{Var}(u_t) = \sigma_e^2 + \alpha_1^2 \sigma_e^2 = \sigma_e^2(1 + \alpha_1^2)$。除非样本中的 x_t 和 x_{t+1} 是不相关的，否则只要 $\alpha \neq 0$，公式中的第二项就是非零的。如果 x_t 和 x_{t+1} 是正相关的，且 $\alpha < 0$，这个真实的方差比通常的方差要小。当方程是水平形式的

（没有经过差分）时，常见的情况是 $\alpha>0$，x_t 和 x_{t+1}之间有正相关。

习题 12.2

$\hat{\rho}\pm1.96\mathrm{se}$（$\hat{\rho}$），其中的 se（$\hat{\rho}$）是回归中所报告的标准误。或者，我们也可以使用异方差—稳健标准误。要证明异方差—稳健标准误是渐近生效的有点复杂，这是因为 OLS 残差取决于 $\hat{\beta}_j$。

习题 12.3

我们想到的模型是 $u_t=\rho_1u_{t-1}+\rho_4u_{t-4}+e_t$。我们想要检验 H_0：$\rho_1=0$，$\rho_4=0$，对立假设是：H_0 是错的。做 $\hat{u}_t$ 对 $\hat{u}_{t-1}$和 $\hat{u}_{t-4}$的回归，求出通常的 F 统计量，检验这两个滞后的联合显著性。（我们是在检验两个约束条件。）

习题 12.4

我们可以用一阶差分估计这个方程。因为 $\hat{\rho}=0.92$ 足够接近于 1，这使得我们对水平形式的回归产生怀疑。更多的讨论参见第 18 章。

习题 12.5 772

因为只有一个解释变量，做怀特检验并不难。只要做 $\hat{u}_t^2$ 对 $return_{t-1}$和 $return_{t-1}^2$（有截距）的回归，并计算 $return_t$ 和 $return_{t-1}^2$的联合显著性 F 检验就行了，如果它们在足够小的显著性水平上是联合显著的，我们就拒绝同方差的虚拟假设。

第 13 章

习题 13.1

是的，假定我们控制了所有的相关因素。*black* 的系数是 1.076，其标准误为 0.174，系数 1.076 在统计上无异于 1。95%的置信区间为 0.735～1.417。

习题 13.2

highearn 的系数表明，在收入封顶值不存在任何变化的情况下，高收入者平均来说在长得多的时间里领取劳工补偿金——约为 29.2%［因为 $\exp(0.256)-1\approx0.292$］。

习题 13.3

首先，$E(v_{i1})=E(a_i+u_{i1})=E(a_i)+E(v_{i1})=0$。类似地，$E(v_{i2})=0$。这样，因为由假定所有的协方差项均为零，$v_{i1}$与 v_{i2}之间的协方差就是 $E(v_{i1}v_{i2})=E[(a_i+u_{i1})(a_i+u_{i2})]=E(a_i^2)+E(a_iu_{i1})+E(a_iu_{i2})+E(u_{i1}u_{i2})=E$

(a_i^2)。但是 $\mathrm{E}(a_i^2)=\mathrm{Var}(a_i)$，因为 $\mathrm{E}(a_i)=0$。这导致各个 i 中误差项存在时间上的正序列相关，这使得混合横截面回归中得出的通常的 OLS 标准误是有偏的。

习题 13.4

因为 $\Delta admn = admn_{90} - admn_{85}$ 是二值标示变量的差分，当且仅当 $admn_{85}=1$ 和 $admn_{90}=0$ 时，它可为 -1。换句话说，华盛顿州 1985 年有一项本身的行政法，但在 1990 年被废止了。

习题 13.5

不，就像它不会在严格的外生解释变量条件下的时间序列回归中得出有偏误、非一致性估计量一样。有两个值得担心的理由：第一，任何方程中的误差项的序列相关一般使通常的 OLS 标准误和检验统计量有偏误。第二，它意味着混合 OLS 不如考虑到序列相关的估计量有效（如第 12 章）。

第 14 章

习题 14.1

无论是否用一阶差分或组内变换，我们估计 $kids_{it}$ 的系数时都将存在问题。例如，运用组内变换，如果 $kids_{it}$ 在家庭 i 之间的变化不大，则 $\ddot{kids}_{it}=kids_{it}-\overline{kids}_i=0(t=1,2,3)$。只要某些家庭在 $kids_{it}$ 上有变异，就可以计算固定效应估计量，但 *kids* 的系数将估计得非常不精确。这是固定效应估计（或一阶差分估计）中多重共线性表现的一个形式。773

习题 14.2

如果一个公司在第一年没有得到许可，在第二年可能得到或得不到许可。但如果它在第一年得到了许可，第二年就不能得到许可。也就是说，如果 $grant_{-1}=1$，则 $grant=0$。这将引起 $grant$ 与 $grant_{-1}$ 之间的负相关。我们可通过用 JTRAIN.RAW 中 1989 年的数据计算一个 $grant$ 对 $grant_{-1}$ 的回归来证实这一点。用样本中的所有公司的数据，得到

$$\widehat{grant}=0.248-0.248grant_{-1}$$
$$(0.035)\quad(0.072)$$
$$n=157, R^2=0.070$$

$grant_{-1}$ 的系数必定为截距的相反数，因为当 $grant_{-1}=1$ 时 $\widehat{grant}=0$。

习题 14.3

这表明未观测到的效应 a_i 与 $union_{it}$ 正相关。记住，混合 OLS 将 a_i 留在

误差项中，然而固定效应消掉了 a_i。根据定义，a_i 对 log（*wage*）有正的效应。通过标准的遗漏变量分析（见第 13 章），当解释变量（*union*）与遗漏变量（a_i）正相关时，OLS 有一个朝上的偏误。这样，隶属一个工会似乎与不随时间变化的、未观测到的影响工资的因素正相关。

习题 14.4

不，如果家庭中所有的姊妹都有相同的父母。那么，因为父母的种族变量在姊妹间不会有变化，在式（14.13）中它们将通过差分而消掉。

第 15 章

习题 15.1

可能不是。在简单方程（15.18）中，受教育年数是误差项的一部分。如果某些被分给小的征兵抽签号的人，获得了更多的学校教育，则抽签号和教育负相关，违背了方程（15.4）中对工具变量的第一个要求。

习题 15.2

（i）为了满足式（15.27），我们要求高中的同类群体效应保持到了大学，即对有给定 *SAT* 分数的学生而言，如果他毕业于吸食大麻越盛行的高中，到了大学就可能吸食越多的大麻。即使识别条件（15.27）成立，这种联系可能也是微弱的。

（ii）我们需要假定高中吸食大麻的学生人数百分比与未观测到的、影响大学平均等级积分点的因素不相关。然而当我们通过将 *SAT* 包括到方程中，稍微控制了高中的质量，这个假定也许是不够的。有可能为学生上大学 774 做了较好准备的高中，学生吸食大麻的也较少。或者，大麻吸食量可能与平均收入水平相关。当然，这都是经验上的问题，我们也许能也许不能给出回答。

习题 15.3

尽管 NRA 和枪支杂志订阅者的盛行很可能与是否存在枪支管理法相关，然而它们与未观测到的、影响暴力犯罪率的因素不相关，这一点并不是显然的。实际上，我们也许认为对枪支感兴趣的人口数是高犯罪率的一个反映，那么控制了经济和人口统计变量是不足以刻画这种关系的。认为这些变量在暴力犯罪方程中是真正外生的，会很难令人信服。

习题 15.4

跟往常一样，有两个要求。第一，在消除掉投资率和劳动力增长的影响后，政府支出增长事实上应当与总统的党派有系统性的关联。换句话说，工

具必须与内生解释变量偏相关。尽管我们可能认为共和党总统情况下政府支出增长得更慢，但无疑这在美国并不总是正确的，可能不得不用诱导型 $gGOV_t = \pi_0 + \pi_1 REP_{t-1} + \pi_2 INVERAT_t + \pi_3 gLAB_t + v_t$ 中 REP_{t-1} 的 t 统计量来进行检验。我们必须假定总统的党派对 $gGDP$ 没有独自的效应。这个假定可能不成立，例如，货币政策因总统党派不同有系统性差异，从而对 GDP 的增长有可分辨的效应。

第 16 章

习题 16.1

可能不。因为企业同时选择价格和广告支出，我们对广告支出做外生变化的实验不感兴趣，我们想知道其对价格的影响。我们倒是应该将价格和广告支出分别建立成需求和成本变量的一个函数。这就离开了经济理论的范畴。

习题 16.2

我们必须做两个假定。第一，货币供给的增长应该出现在方程（16.22）中，所以它与 *inf* 偏相关。第二，我们必须假定货币供给的增长不在方程（16.23）中出现。如果我们认为必须在方程（16.23）中包括货币供给增长，那我们仍缺少 *inf* 的一个工具变量。当然，假定货币供给增长是外生的也有问题。

习题 16.3

利用第 15 章中的 Hausman 检验。具体而言，令 $\hat{v}_2$ 表示 *open* 对 log（*pcinc*）和 log（*land*）的约简型回归中的 OLS 残差。然后，利用 *inf* 对 *open*，log（*pcinc*）和 $\hat{v}_2$ 的一个 OLS 回归，并计算 $\hat{v}_2$ 的显著性的 t 统计量。
775 如果 $\hat{v}_2$ 显著，则 2SLS 和 OLS 估计值在统计上是不同的。

习题 16.4

需求方程看起来像

$$\log(fish_t) = \beta_0 + \beta_1 \log(prcfish_t) + \beta_2 \log(inc_t) + \beta_3 \log(prcchick_t) + \beta_4 \log(prcbeef_t) + u_{t1}$$

式中，使用对数是为了使所有的弹性都是常数。根据假定，需求函数不包含季节性，所以方程不包含月度虚拟变量（如以 1 月份为基月的 feb_t，mar_t，…，dec_t）。同样，根据假定，鱼的供给有季节性，即供给函数至少取决于某些月度虚拟变量。即便不求出 log（*prcfish*）的约简型，我们也能断定它取决于月度虚拟变量。由于这些都是外生变量，所以它们都能用做需求方程

中 log（*prcfish*）的工具变量。因此，我们可以利用月度虚拟变量作为 log（*prcfish*）的工具变量而估计鱼的需求方程。识别性则要求至少有一个月度虚拟变量在 log（*prcfish*）的约简型中有非零的系数。

第 17 章

习题 17.1

$H_0: \beta_4 = \beta_5 = \beta_6 = 0$，所以有 3 个约束，因此在 *LR* 或瓦尔德检验中有 3 个自由度。

习题 17.2

我们需要 $\Phi(\hat{\beta}_0 + \hat{\beta}_1 nwifeinc + \hat{\beta}_2 educ + \hat{\beta}_3 exper + \hat{\beta}_4 exper^2 + \cdots)$对 *exper* 的偏导数 $\phi(\cdot)(\hat{\beta}_3 - 2\hat{\beta}_4 exper)$，其中 $\phi(\cdot)$是在给定变量值和工作经历的初始水平上计算出来的。因此，我们需要计算标准正态概率分布的值 $0.270 - 0.012(20.13) + 0.131(12.3) + 0.123(10) - 0.001\,9(10^2) - 0.053(42.5)$ $0.868(0) + 0.036(1) \approx 0.463$，其中我们代入了工作经历的初始值 10。但 $\phi(0.463) = (2\pi)^{-1/2}\exp[-(0.463^2)/2] \approx 0.358$。接下来，我们将它乘以 $\hat{\beta}_3 + 2\hat{\beta}_4 exper$ 在 $exper = 10$ 处的值。利用微积分近似的偏效应是 $0.358[0.123 - 2(0.001\,9)(10)] \approx 0.030$。换句话说，在给定解释变量值和 $exper = 10$ 的起点处，下一年的工作经历使参与劳动市场的概率提高约 0.03。

习题 17.3

不是。婚外情次数是一个非负整数，可能总体中相当大的比例都取值零或很小的数字。Tobit 模型尽管容许有大量的零，但仍将 y 在正值上的分布看成连续分布。所以在这里用 Tobit 模型并不现实。假定$y = \max(0, y^*)$(其中 y^* 正态分布)与婚外情在 $y > 0$ 时的离散性完全不相称。

776 习题 17.4

调整标准误等于通常的泊松 MLE 标准误乘以 $\hat{\sigma} = \sqrt{2} \approx 1.41$，所以调整标准误约高出 41%。准 *LR* 统计量等于通常的 *LR* 统计量除以$\hat{\sigma}^2$，所以它将是通常 *LR* 统计量的一半。

习题 17.5

根据假定，$mvp_i = \beta_0 + \mathbf{x}_i\boldsymbol{\beta} + u_i$，其中 $\mathbf{x}_i\boldsymbol{\beta}$ 和往常一样表示外生变量的一个线性函数。现在，所观测到的工资是最低工资和边际价值产品中的最大者，即$wage_i = \max(minwage_i, mvp_i)$，这与方程（17.34）十分类似，区别只是用最大算子取代了最小算子。

第 18 章

习题 18.1

可以把这些值直接代入式（18.1），并取其期望值。首先，因为 $s<0$ 时 $z_s=0$，于是 $y_{-1}=\alpha+u_{-1}$。又因为 $z_0=1$，所以，$y_0=\alpha+\delta_0+u_0$。当 $h\geqslant 1$ 时，$y_h=\alpha+\delta_{h-1}+\delta_h+u_h$。因为误差的期望值为零，所以当 $h\geqslant 1$ 时有 $\mathrm{E}(y_{-1})=\alpha$，$\mathrm{E}(y_0)=\alpha+\delta_0$ 及 $\mathrm{E}(y_h)=\alpha+\delta_{h-1}+\delta_h$。随着 h 趋于无穷，$\delta_h\rightarrow 0$。于是，h 趋于无穷时 $\mathrm{E}(y_h)\rightarrow\alpha$，也就是说，$y_h$ 的期望值趋近于它在时间 0（即 z 变大）之前的期望值。它的含义是：z 的增长持续了两个时期，这个增长只是暂时性的。

习题 18.2

根据所描述的构造方法，Δy_t 和 Δx_t 是 $i.i.d$ 序列，且相互独立。当然，它们肯定是不相关的。若 $\hat{\gamma}_1$ 是 Δy_t 对 Δx_t（$t=1,2,\cdots,n$）的回归中的斜率系数，就有 $\operatorname{plim}\hat{\gamma}_1=0$。这个结果是必然的，因为我们是在做一个 I（0）过程对另一个 I（0）过程的回归，而它们之间又是不相关的。我们把方程写为 $\Delta y_t=\gamma_0+\gamma_1\Delta x_t+e_t$，其中，$\gamma_0=\gamma_1=0$。因为 $\{e_t\}$ 是独立于 $\{\Delta x_t\}$ 的，严格外生性假定成立。另外，$\{e_t\}$ 是序列不相关和同方差的。根据第 11 章的定理 11.2，$\hat{\gamma}_1$ 的 t 统计量有渐近标准正态分布。如果 e_t 是正态分布的，经典线性模型假定成立，这个 t 统计量就恰好有 t 分布。

习题 18.3

令 $x_t=x_{t-1}+a_t$，其中的 $\{a_t\}$ 是 I（0）。根据假定，存在着某种线性组合，比如说 $s_t=y_t-\beta x_t$，它是个 I（0）。于是，$y_t-\beta x_t=y_t-\beta(x_{t-1}+a_t)=s_t+\beta a_t$。因为依假定 s_t 和 a_t 是 I（0），所以 $s_t+\beta a_t$ 也是 I（0）。

习题 18.4

使用残差平方和形式的 F 检验，并假定同方差性，受约束的 SSR 可以通过做 $\Delta hy6_t-\Delta hy3_{t-1}+(hy6_{t-1}-hy3_{t-2})$ 对一个常数的回归得到。当施加约束条件时，α_0 是方程 $\Delta hy6_t=\alpha_0+\gamma_0\Delta hy3_{t-1}+\delta(hy6_{t-1}-hy3_{t-2})$ 中惟一需要估计的参数。不受约束的残差平方和可以从方程（18.39）中求出来。

777

习题 18.5

我们在这里要拟合两个方程 $\hat{y}_t=\hat{\alpha}+\hat{\beta}t$ 和 $\hat{y}_t=\hat{\gamma}+\hat{\delta}year_t$。由于 $year_t=t+49$，我们可以求出参数之间的关系：把它代入第二个方程得到 $\hat{y}_t=\hat{\gamma}+\hat{\delta}(t+49)=(\hat{\gamma}+49\hat{\delta})+\hat{\delta}t$。对比新方程和第一个方程中的斜率和截距，得到 $\hat{\delta}=\hat{\beta}$——以使 $year_t$ 和 t 的斜率相等——及 $\hat{\alpha}=\hat{\gamma}+49\hat{\delta}$。一般情况下，

如果我们用 *year* 而不用 t，最多会使截距有所改变，但斜率却是一样的。(你可以利用时间序列数据，比如 HSEINV.RAW 或 INVEN.RAW，来验证这一点。）无论使用 t 还是使用其他表示时间的指标，都不会改变拟合值的大小，因此，当然也不会改变将来值的预测结果。至于截距，则只要根据在回归中添加趋势的方法的不同而适当地调整一下就可以了。

附录 G　统计学用表

表 G.1　标准正态分布下的累积面积

z	0	1	2	3	4	5	6	7	8	9
−3.0	0.001 3	0.001 3	0.001 3	0.001 2	0.001 2	0.001 1	0.001 1	0.001 1	0.001 0	0.001 0
−2.9	0.001 9	0.001 8	0.001 8	0.001 7	0.001 6	0.001 6	0.001 5	0.001 5	0.001 4	0.001 4
−2.8	0.002 6	0.002 5	0.002 4	0.002 3	0.002 3	0.002 2	0.002 1	0.002 1	0.002 0	0.001 9
−2.7	0.003 5	0.003 4	0.003 3	0.003 2	0.003 1	0.003 0	0.002 9	0.002 8	0.002 7	0.002 6
−2.6	0.004 7	0.004 5	0.004 4	0.004 3	0.004 1	0.004 0	0.003 9	0.003 8	0.003 7	0.003 6
−2.5	0.006 2	0.006 0	0.005 9	0.005 7	0.005 5	0.005 4	0.005 2	0.005 1	0.004 9	0.004 8
−2.4	0.008 2	0.008 0	0.007 8	0.007 5	0.007 3	0.007 1	0.006 9	0.006 8	0.006 6	0.006 4
−2.3	0.010 7	0.010 4	0.010 2	0.009 9	0.009 6	0.009 4	0.009 1	0.008 9	0.008 7	0.008 4
−2.2	0.013 9	0.013 6	0.013 2	0.012 9	0.012 5	0.012 2	0.011 9	0.011 6	0.011 3	0.011 0
−2.1	0.017 9	0.017 4	0.017 0	0.016 6	0.016 2	0.015 8	0.015 4	0.015 0	0.014 6	0.014 3
−2.0	0.022 8	0.022 2	0.021 7	0.021 2	0.020 7	0.020 2	0.019 7	0.019 2	0.018 8	0.018 3
−1.9	0.028 7	0.028 1	0.027 4	0.026 8	0.026 2	0.025 6	0.025 0	0.024 4	0.023 9	0.023 3
−1.8	0.035 9	0.035 1	0.034 4	0.033 6	0.032 9	0.032 2	0.031 4	0.030 7	0.030 1	0.029 4

续前表

z	0	1	2	3	4	5	6	7	8	9
1.7	0.044 6	0.043 6	0.042 7	0.041 8	0.040 9	0.040 1	0.039 2	0.038 4	0.037 5	0.036 7
1.6	0.054 8	0.053 7	0.052 6	0.051 6	0.050 5	0.049 5	0.048 5	0.047 5	0.046 5	0.045 5
−1.5	0.066 8	0.065 5	0.064 3	0.063 0	0.061 8	0.060 6	0.059 4	0.058 2	0.057 1	0.055 9
−1.4	0.080 8	0.079 3	0.077 8	0.076 4	0.074 9	0.073 5	0.072 1	0.070 8	0.069 4	0.068 1
−1.3	0.096 8	0.095 1	0.093 4	0.091 8	0.090 1	0.088 5	0.086 9	0.085 3	0.083 8	0.082 3
−1.2	0.115 1	0.113 1	0.111 2	0.109 3	0.107 5	0.105 6	0.103 8	0.102 0	0.100 3	0.098 5
−1.1	0.135 7	0.133 5	0.131 4	0.129 2	0.127 1	0.125 1	0.123 0	0.121 0	0.119 0	0.117 0
1.0	0.158 7	0.156 2	0.153 9	0.151 5	0.149 2	0.146 9	0.144 6	0.142 3	0.140 1	0.137 9
−0.9	0.184 1	0.181 4	0.178 8	0.176 2	0.173 6	0.171 1	0.168 5	0.166 0	0.163 5	0.161 1
−0.8	0.211 9	0.209 0	0.206 1	0.203 3	0.200 5	0.197 7	0.194 9	0.192 2	0.189 4	0.186 7
−0.7	0.242 0	0.238 9	0.235 8	0.232 7	0.229 6	0.226 6	0.223 6	0.220 6	0.217 7	0.214 8
0.6	0.274 3	0.270 9	0.267 6	0.264 3	0.261 1	0.257 8	0.254 6	0.251 4	0.248 3	0.245 1
0.5	0.308 5	0.305 0	0.301 5	0.298 1	0.294 6	0.291 2	0.287 7	0.284 3	0.281 0	0.277 6
−0.4	0.344 6	0.340 9	0.337 2	0.333 6	0.330 0	0.326 4	0.322 8	0.319 2	0.315 6	0.312 1
0.3	0.382 1	0.378 3	0.374 5	0.370 7	0.366 9	0.363 2	0.359 4	0.355 7	0.352 0	0.348 3
−0.2	0.420 7	0.416 8	0.412 9	0.409 0	0.405 2	0.401 3	0.397 4	0.393 6	0.389 7	0.385 9
−0.1	0.460 2	0.456 2	0.452 2	0.448 3	0.444 3	0.440 4	0.436 4	0.432 5	0.428 6	0.424 7
−0.0	0.500 0	0.496 0	0.492 0	0.488 0	0.484 0	0.480 1	0.476 1	0.472 1	0.468 1	0.464 1
0.0	0.500 0	0.504 0	0.508 0	0.512 0	0.516 0	0.519 9	0.523 9	0.527 9	0.531 9	0.535 9
0.1	0.539 8	0.543 8	0.547 8	0.551 7	0.555 7	0.559 6	0.563 6	0.567 5	0.571 4	0.575 3
0.2	0.579 3	0.583 2	0.587 1	0.591 0	0.594 8	0.598 7	0.602 6	0.606 4	0.610 3	0.614 1
0.3	0.617 9	0.621 7	0.625 5	0.629 3	0.633 1	0.636 8	0.640 6	0.644 3	0.648 0	0.651 7
0.4	0.655 4	0.659 1	0.662 8	0.666 4	0.670 0	0.673 6	0.677 2	0.680 8	0.684 4	0.687 9
0.5	0.691 5	0.695 0	0.698 5	0.701 9	0.705 4	0.708 8	0.712 3	0.715 7	0.719 0	0.722 4
0.6	0.725 7	0.729 1	0.732 4	0.735 7	0.738 9	0.742 2	0.745 4	0.748 6	0.751 7	0.754 9
0.7	0.758 0	0.761 1	0.764 2	0.767 3	0.770 4	0.773 4	0.776 4	0.779 4	0.782 3	0.785 2
0.8	0.788 1	0.791 0	0.793 9	0.796 7	0.799 5	0.802 3	0.805 1	0.807 8	0.810 6	0.813 3
0.9	0.815 9	0.818 6	0.821 2	0.823 8	0.826 4	0.828 9	0.831 5	0.834 0	0.836 5	0.838 9
1.0	0.841 3	0.843 8	0.846 1	0.848 5	0.850 8	0.853 1	0.855 4	0.857 7	0.859 9	0.862 1
1.1	0.864 3	0.866 5	0.868 6	0.870 8	0.872 9	0.874 9	0.877 0	0.879 0	0.881 0	0.883 0
1.2	0.884 9	0.886 9	0.888 8	0.890 7	0.892 5	0.894 4	0.896 2	0.898 0	0.899 7	0.901 5
1.3	0.903 2	0.904 9	0.906 6	0.908 2	0.909 9	0.911 5	0.913 1	0.914 7	0.916 2	0.917 7
1.4	0.919 2	0.920 7	0.922 2	0.923 6	0.925 1	0.926 5	0.927 9	0.929 2	0.930 6	0.931 9

续前表

z	0	1	2	3	4	5	6	7	8	9
1.5	0.933 2	0.934 5	0.935 7	0.937 0	0.938 2	0.939 4	0.940 6	0.941 8	0.942 9	0.944 1
1.6	0.945 2	0.946 3	0.947 4	0.948 4	0.949 5	0.950 5	0.951 5	0.952 5	0.953 5	0.954 5
1.7	0.955 4	0.956 4	0.957 3	0.958 2	0.959 1	0.959 9	0.960 8	0.961 6	0.962 5	0.963 3
1.8	0.964 1	0.964 9	0.965 6	0.966 4	0.967 1	0.967 8	0.968 6	0.969 3	0.969 9	0.970 6
1.9	0.971 3	0.971 9	0.972 6	0.973 2	0.973 8	0.974 4	0.975 0	0.975 6	0.976 1	0.976 7
2.0	0.977 2	0.977 8	0.978 3	0.978 8	0.979 3	0.979 8	0.980 3	0.980 8	0.981 2	0.981 7
2.1	0.982 1	0.982 6	0.983 0	0.983 4	0.983 8	0.984 2	0.984 6	0.985 0	0.985 4	0.985 7
2.2	0.986 1	0.986 4	0.986 8	0.987 1	0.987 5	0.987 8	0.988 1	0.988 4	0.988 7	0.989 0
2.3	0.989 3	0.989 6	0.989 8	0.990 1	0.990 4	0.990 6	0.990 9	0.991 1	0.991 3	0.991 6
2.4	0.991 8	0.992 0	0.992 2	0.992 5	0.992 7	0.992 9	0.993 1	0.993 2	0.993 4	0.993 6
2.5	0.993 8	0.994 0	0.994 1	0.994 3	0.994 5	0.994 6	0.994 8	0.994 9	0.995 1	0.995 2
2.6	0.995 3	0.995 5	0.995 6	0.995 7	0.995 9	0.996 0	0.996 1	0.996 2	0.996 3	0.996 4
2.7	0.996 5	0.996 6	0.996 7	0.996 8	0.996 9	0.997 0	0.997 1	0.997 2	0.997 3	0.997 4
2.8	0.997 4	0.997 5	0.997 6	0.997 7	0.997 7	0.997 8	0.997 9	0.997 9	0.998 0	0.998 1
2.9	0.998 1	0.998 2	0.998 2	0.998 3	0.998 4	0.998 4	0.998 5	0.998 5	0.998 6	0.998 6
3.0	0.998 7	0.998 7	0.998 7	0.998 8	0.998 8	0.998 9	0.998 9	0.998 9	0.999 0	0.999 0

例：若 $Z\sim\text{Normal}(0,1)$，则 $P(Z\leqslant-1.32)=0.093\,4$ 和 $P(Z\leqslant 1.84)=0.967\,1$。

资料来源：本表由 stata$^{®}_{\text{function normal}}$ 生成。

表 G.2 *t* 分布的临界值

		显著性水平				
	单尾：	0.10	0.05	0.025	0.01	0.005
	双尾：	0.20	0.10	0.05	0.02	0.01
自由度	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	6.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.067	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

例：自由度为25的单尾检验的1%临界值是2.485。大（>120）自由度双尾检验的5%临界值是1.96。

资料来源：本表由stata® function normd生成。

表 G.3a F 分布的 10%临界值

		分子自由度									
		1	2	3	4	5	6	7	8	9	10
分母自由度	10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
	11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
	12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
	13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
	14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
	15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
	16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
	17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
	18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
	19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
	20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
	21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
	22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
	23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
	24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
	25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
	26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.83
	27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
	28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
	29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
	30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
	40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
	60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
	90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
	120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
	∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

例：分子自由度为 $df=2$ 而分母自由度 $df=40$ 的 10%临界值是 2.44。

资料来源：本表由 stata®function normd 生成。

表 G.3b **F 分布的 5%临界值**

		分子自由度									
		1	2	3	4	5	6	7	8	9	10
分母自由度	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
	14	4.60	3.74	3.34	3.11	2.93	2.85	2.76	2.70	2.65	2.60
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
	90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

例：分子自由度为 $df=4$ 而分母自由度 $df(\infty)$的 5%临界值是 2.37。

资料来源：本表由 stata® function normd 生成。

表 G.3c F 分布的 1%临界值

分母自由度	分子自由度 1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.31

例：分子自由度为 $df-3$ 而分母自由度 $df=60$ 的 1%临界值是 4.13。

资料来源：本表由 stata® function normd 生成。

表 G.4　　χ平方分布的临界值

		显著性水平		
		0.10	**0.05**	**0.01**
自由度	1	2.71	3.84	6.63
	2	4.61	5.99	9.21
	3	6.25	7.81	11.34
	4	7.78	9.49	13.28
	5	9.24	11.07	15.09
	6	10.64	12.59	16.81
	7	12.02	14.07	18.48
	8	13.36	15.51	20.09
	9	14.68	16.92	21.67
	10	15.99	18.31	23.21
	11	17.28	19.68	24.72
	12	18.55	21.03	26.22
	13	19.81	22.36	27.69
	14	21.06	23.68	29.14
	15	22.31	25.00	30.58
	16	23.54	26.30	32.00
	17	24.77	27.59	33.41
	18	25.99	28.87	34.81
	19	27.20	30.14	36.19
	20	28.41	31.41	37.57
	21	29.62	32.67	38.93
	22	30.81	33.92	40.29
	23	32.01	35.17	41.64
	24	33.20	36.42	42.98
	25	34.38	37.65	44.31
	26	35.56	38.89	45.64
	27	36.74	40.11	46.96
	28	37.92	41.34	48.28
	29	39.09	42.56	49.59
	30	40.26	43.77	50.89

例：自由度为 $df=8$ 的 5%临界值是 15.51。

资料来源：本表由 stata® function normd 生成。

参考文献

Angrist, J. D. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review* 80,313 – 336.

Angrist, J. D., and A. B. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106, 979 – 1014.

Ashenfelter, O., and A. B. Krueger (1994), "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review* 84, 1157 – 1173.

Averett, S., and S. Korenman (1996), "The Economic Reality of the Beauty Myth," *Journal of Human Resources* 31, 304 – 330.

Ayers, I., and S. D. Levitt (1998), "Measuring Positive Externalities from Unobservable Victim Precaution: An Empirical Analysis of Lojack," *Quarterly Journal of Economics* 108, 43 – 77.

Banerjee, A., J. Dolado, J. W. Galbraith, and D. F. Hendry (1993), *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Bartik, T. J. (1991), "The Effects of Property Taxes and Other Local Public Policies on the Intrametropolitan Pattern of Business Location," in *Industry Location and Public Policy*. Ed. H. W. Herzog and A. M. Schlottmann, 57 – 80. Knoxville: University of Tennessee Press.

Becker, G. S. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76, 169 – 217.

Belsley, D., E. Kuh, and R. Welsch (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Berk, R. A. (1990), "A Primer on Robust Regression," in *Modern Methods of Data Analysis*. Ed. J. Fox and J. S. Long, 292 – 324. Newbury Park, CA: Sage Publications.

Betts, J. R. (1995), "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics* 77, 231 – 250.

Biddle, J. E., and D. S. Hamermesh (1990), "Sleep and the Allocation of Time," *Journal of Political Economy* 98, 922 – 943.

Biddle, J. E., and D. S. Hamermesh (1998), "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics* 16, 172 – 201.

Blackburn, M., and S. Korenman (1994), "The Declining Marital-Status Earnings Differential," *Journal of Population Economics* 7, 247 – 270.

Blackburn, M., and D. Neumark (1992), "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics* 107, 1421 – 1436.

Blömstrom, M., R. E. Lipsey, and M. Zejan (1996), "Is Fixed Investment the Key to Economic Growth?" *Quarterly Journal of Economics* 111, 269 – 276.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52, 5 – 59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994), "ARCH Models," Chapter 49 in *Handbook of Econometrics*, Volume 4. Ed. R. F. Engle and D. L. McFadden, 2959 – 3038. Amsterdam: North-Holland.

Bound, J., D. A. Jaeger, and R. M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association* 90, 443 – 450.

Breusch, T. S., and A. R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica* 50, 987 – 1007.

Cameron, A. C., and P. K. Trivedi (1998), *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Campbell, J. Y., and N. G. Mankiw (1990), "Permanent Income, Current Income, and Consumption," *Journal of Business and Economic Statistics* 8, 265 – 279.

Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*. Ed. L. N. Christophides, E. K. Grant, and R. Swidinsky, 201 – 222. Toronto: University of Toronto Press.

Card, D., and A. Krueger (1992), "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100, 1 – 40.

Castillo-Freeman, A. J., and R. B. Freeman (1992), "When the Minimum Wage Really Bites: The Effect of the U. S.-Level Minimum on Puerto Rico," in *Immigration and the Work Force*. Ed. G. J. Borjas and R. B. Freeman, 177 – 211. Chicago: University of Chicago Press.

Clark, K. B. (1984), "Unionization and Firm Performance: The Impact on Profits, Growth, and Productivity," *American Economic Review* 74, 893 – 919.

Cloninger, D. O. (1991), "Lethal Police Response as a Crime Deterrent: 57-City Study Suggests a Decrease in Certain Crimes," *American Journal of Economics and Sociology* 50, 59 - 69.

Cloninger, D. O., and L. C. Sartorius (1979), "Crime Rates, Clearance Rates and Enforcement Effort: The Case of Houston, Texas," *American Journal of Economics and Sociology* 38, 389 - 402.

Cochrane, J. H. (1997), "Where is the Market Going? Uncertain Facts and Novel Theories," *Economic Perspectives* 21, Federal Reserve Bank of Chicago, 3 - 37.

Cornwell, C., and W. N. Trumbull (1994), "Estimating the Economic Model of Crime Using Panel Data," *Review of Economics and Statistics* 76, 360 - 366.

Currie, J. (1995), *Welfare and the Well-Being of Children*. Chur, Switzerland: Harwood Academic Publishers.

Currie, J., and N. Cole (1993), "Welfare and Child Health: The Link Between AFDC Participation and Birth Weight," *American Economic Review* 83, 971 - 983.

Currie, J., and D. Thomas (1995), "Does Head Start Make a Difference?" *American Economic Review* 85, 341 - 364.

Davidson, R., and J. G. MacKinnon (1981), "Several Tests of Model Specification in the Presence of Alternative Hypotheses," *Econometrica* 49, 781 - 793.

Davidson, R., and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*. New York: Oxford University Press.

De Long, J. B., and L. H. Summers (1991), "Equipment Investment and Economic Growth," *Quarterly Journal of Economics* 106, 445 - 502.

Dickey, D. A., and W. A. Fuller (1979), "Distributions of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427 - 431.

Diebold, F. X. (1998), *Elements of Forecasting*. Cincinnati, OH: South-Western.

Downes, T. A., and S. M. Greenstein (1996), "Understanding the Supply Decisions of Nonprofits: Modeling the Location of Private Schools," *Rand Journal of Economics* 27, 365 - 390.

Draper, N., and H. Smith (1981), *Applied Regression Analysis*. 2d ed. New York: Wiley.

Durbin, J. (1970), "Testing for Serial Correlation in Least Squares Regressions When Some of the Regressors are Lagged Dependent Variables," *Econometrica* 38, 410 - 421.

Durbin, J., and G. S. Watson (1950), "Testing for Serial Correlation in Least Squares Regressions I," *Biometrika* 37, 409 - 428.

Eicker, F. (1967), "Limit Theorems for Regressions with Unequal and Dependent Errors," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 59 - 82. Berkeley: University of California Press.

Eide, E. (1994), *Economics of Crime: Deterrence and the Rational Offender*. Amsterdam: North Holland.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50, 987 - 1007.

Engle, R. F., and C. W. J. Granger (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, 251 - 276.

Evans, W. N., and R. M. Schwab (1995), "Finishing High School and Starting College:

Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics* 110, 941-974.

Fair, R. C. (1996), "Econometrics and Presidential Elections," *Journal of Economic Perspectives* 10, 89-102.

Friedman, B. M., and K. N. Kuttner (1992), "Money, Income, Prices, and Interest Rates," *American Economic Review* 82, 472 492.

Garen, J. E (1994), "Executive Compensation and Principal-Agent Theory," *Journal of Political Economy* 102, 175-1199.

Geronimus, A. T., and S. Korenman (1992), "The Socioeconomic Consequences of Teen Childbearing Reconsidered," *Quarterly Journal of Economics* 107, 1187-1214.

Goldberger, A. S. (1991), *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Granger, C. W. J., and P. Newbold (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics* 2, 111-120.

Greene, W. (1997), *Econometric Analysis*. 3rd edition New York: MacMillan.

Griliches, Z. (1957), "Specification Bias in Estimates of Production Functions," *Journal of Farm Economics* 39,8-20.

Grogger, J. (1990), "The Deterrent Effect of Capital Punishment: An Analysis of Daily Homicide Counts," *Journal of the American Statistical Association* 410, 295-303.

Grogger, J. (1991), "Certainty vs. Severity of Punishment," *Economic Inquiry* 29, 297-309.

Hall, R. J. (1988), "The Relation Between Price and Marginal Cost in U. S. Industry," *Journal of Political Economy* 96, 921-948.

Hamermesh, D. S., and J. E. Biddle (1994), "Beauty and the Labor Market," *American Economic Review* 84, 1174-1194.

Hamilton, J. D. (1994), *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hanushek, E. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 1141-1177.

Harvey, A. (1990), *The Econometric Analysis of Economic Time Series*. 2d ed. Cambridge, MA: MIT Press.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica* 46, 1251-1271.

Hausman, J. A., and D. A. Wise (1977), "Social Experimentation, Truncated Distributions, and Efficient Estimation," *Econometrica* 45, 319-339.

Herrnstein, R. J., and C. Murray (1994), *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hersch, J., and L. S. Stratton (1997), "Housework, Fixed Effects, and Wages of Married Workers," *Journal of Human Resources* 32, 285-307.

Hines, J. R. (1996), "Altered States: Taxes and the Location of Foreign Direct Investment in America," *American Economic Review* 86, 1076-1094.

Holzer, H. (1991), "The Spatial Mismatch Hypothesis: What Has the Evidence Shown?" *Urban Studies* 28, 105-122.

Holzer, H., R. Block, M. Cheatham, and J. Knott (1993), "Are Training Subsidies Effective? The Michigan Experience," *Industrial and Labor Relations Review* 46, 625-636.

Hoxby, C. M. (1994), "Do Private Schools Provide Competition for Public Schools?" National Bureau of Economic Research Working Paper Number 4978.

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 221 – 233. Berkeley: University of California Press.

Hunter, W. C., and M. B. Walker (1996), "The Cultural Affinity Hypothesis and Mortgage Lending Decisions," *Journal of Real Estate Finance and Economics* 13, 57 – 70.

Hylleberg, S. (1991), *Modelling Seasonality*. Oxford: Oxford University Press.

Kane, T. J., and C. E. Rouse (1995), "Labor-Market Returns to Two- and Four-Year Colleges," *American Economic Review* 85, 600 – 614.

Kiel, K. A., and K. T. McClain (1995), "House Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation," *Journal of Environmental Economics and Management* 28, 241 – 255.

Kleck, G., and E. B. Patterson (1993), "The Impact of Gun Control Ownership Levels on Violence Rates," *Journal of Quantitative Criminology* 9, 249 – 287.

Koenker, R. (1981), "A Note on Studentizing a Test for Heteroskedasticity," *Journal of Econometrics* 17, 107 – 112.

Korenman, S., and D. Neumark (1991), "Does Marriage Really Make Men More Productive?" *Journal of Human Resources* 26, 282 – 307.

Korenman, S., and D. Neumark (1992), "Marriage, Motherhood, and Wages," *Journal of Human Resources* 27, 233 – 255.

Krueger, A. B. (1993), "How Computers Have Changed the Wage Structure: Evidence from Microdata. 1984 – 1989." *Quarterly Journal of Economics* 108, 33 – 60.

Krupp, C. M., and P. S. Pollard (1996), "Market Responses to Antidumping Laws: Some Evidence from the U. S. Chemical Industry," *Canadian Journal of Economics* 29, 199 – 227.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54, 159 – 178.

Larsen, R. J., and M. L. Marx(1986), *An Introduction to Mathematical Statistics and Its Applications*. 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.

Leamer, E. E. (1983.), "Let's Take the Con Out of Econometrics," *American Economic Review* 73, 31 – 43.

Levine, P. B., A. B. Trainor, and D. J. Zimmerman (1996), "The Effect of Medicaid Abortion Funding Restrictions on Abortions, Pregnancies, and Births," *Journal of Health Economics* 15, 555 – 578.

Levine, P. B., and D. J. Zimmerman (1995), "The Benefit of Additional High-School Math and Science Classes for Young Men and Women," *Journal of Business and Economics Statistics* 13, 137 – 149.

Levitt, S. D. (1994), "Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U. S. House," *Journal of Political Economy* 102, 777 – 798.

Levitt, S. D. (1996), "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Legislation," *Quarterly Journal of Economics* 111, 319 – 351.

Low, S. A., and L. R. McPheters (1983), "Wage Differentials and the Risk of Death:

An Empirical Analysis," *Economic Inquiry* 21, 271-280.

Lynch, L. M. (1992), "Private Sector Training and the Earnings of Young Workers," *American Economic Review* 82, 299-312.

MacKinnon, J. G., and H. White (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics* 29, 305-325.

Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maloney, M. T., and R. E. McCormick (1993), "An Examination of the Role that Intercollegiate Athletic Participation Plays in Academic Achievement: Athletes'Feats in the Classroom," *Journal of Human Resources* 28, 555-570.

Mankiw, N. G. (1994), *Macroeconomics*. 2d ed. New York: Worth.

McCarthy, P. S. (1994), "Relaxed Speed Limits and Highway Safety: New Evidence From California," *Economics Letters* 46, 173-179.

McClain, K. T., and J. M. Wooldridge (1995), "A Simple Test for the Consistency of Dynamic Linear Regression in Rational Distributed Lag Models," *Economics Letters* 48, 235-240.

McCormick, R. E., and M. Tinsley (1987), "Athletics versus Academics: Evidence from SAT Scores," *Journal of Political Economy* 95, 1103-1116.

McFadden, D. L. (1974), "Conditional Logit Analysis of Qualitative Choice Analysis," in *Frontiers in Econometrics*. Ed. P. Zarembka, 105-142. New York: Academic Press.

Meyer, B. D. (1995), "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics* 13, 151-161.

Meyer, B. D., W. K. Viscusi, and D. L. Durbin (1995), "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment," *American Economic Review* 85, 322-340.

Mizon, G. E., and J. F. Richard (1986), "The Encompassing Principle and Its Application to Testing Nonnested Hypotheses," *Econometrica* 54, 657-678.

Mroz, T. A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica* 55, 765-799.

Mullahy, J., and P. R. Portney (1990), "Air Pollution, Cigarette Smoking, and the Production of Respiratory Health," *Journal of Health Economics* 9, 193-205.

Mullahy, J., and J. L. Sindelar (1994), "Do Drinkers Know When to Say When? An Empirical Analysis of Drunk Driving," *Economic Inquiry* 32, 383-394.

Netzer, D. (1992), "Differences in Reliance on User Charges by American State and Local Governments," *Public Finance Quarterly* 20, 499-511.

Neumark, D., and W. Wascher (1995), "Minimum Wage Effects on Employment and School Enrollment," *Journal of Business and Economic Statistics* 13, 199-206.

Newey, W. K., and K. D. West (1987), "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Papke, L. E. (1987), "Subnational Taxation and Capital Mobility: Estimates of Tax-Price Elasticities," *National Tax Journal* 40, 191-203.

Papke, L. E. (1994), "Tax Policy and Urban Development: Evidence from the Indiana Enterprise Zone Program," *Journal of Public Economics* 54, 37 – 49.

Papke, L.E. (1995), "Participation in and Contributions to 401(k) Pension Plans: Evidence from Plan Data," *Journal of Human Resources* 30, 311 – 325.

Papke, L.E. (1999), "Are 401(k) Plans Replacing Other Employer-Provided Pensions? Evidence from Panel Data," *Journal of Human Resources*, 34, 346 - 368.

Park R. (1966), "Estimation with Heteroskedastic Error Terms," *Econometrica* 34, 888.

Pavlik, E.L., and A. Belkaoui (1991), *Determinants of Executive Compensation*. New York: Quorum Books.

Peek, J. (1982), "Interest Rates, Income Taxes, and Anticipated Inflation," *American Economic Review* 72, 980 – 991.

Pindyck, R.S., and D.L. Rubinfeld (1992), *Microeconomics*. 2d ed. New York: MacMillan.

Ram, R. (1986), "Government Size and Economic Growth: A New Framework and Some Evidence from Cross-Section and Time-Series Data," *American Economics Review* 76, 191 – 203.

Ramanathan, R (1995), *Introductory Econometrics with Applications*. 3d ed. Fort Worth: Dryden Press.

Ramey, V. (1991), "Nonconvex Costs and the Behavior of Inventories," *Journal of Political Economy* 99, 306 334.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Analysis," *Journal of the Royal Statistical Association*. Series B, 71, 350 – 371.

Romer, D. (1993), "Openness and Inflation: Theory and Evidence," *Quarterly Journal of Economics* 108, 869 – 903.

Rose, N.L. (1985), "The Incidence of Regulatory Rents in the Motor Carrier Industry," *Rand Journal of Economics* 16, 299 318.

Rose, N.L., and A. Shepard (1997), "Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment?" *Rand Journal of Economics* 28, 489 – 514.

Rouse, C.E. (1998), "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *Quarterly Journal of Economics* 113, 553 – 602.

Sander, W. (1992), "The Effect of Women's Schooling on Fertility," *Economic Letters* 40, 229 - 233.

Savin, N. E., and K.J. White (1977), "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors," *Econometrica* 45, 1989 - 1996.

Shea, J. (1993), "The Input-Output Aproach to Instrument Selection," *Journal of Business and Economic Statistics* 11, 145 – 155.

Shugart, W.F., and R.D. Tollison (1984), "The Random Character of Merger Activity," *Rand Journal of Economics* 15, 500 – 509.

Solon, G. (1985), "The Minimum Wage and Teen-age Employment: A Re-analysis with Attention to Serial Correlation and Seasonality," *Journal of Human Resources* 20, 292 – 297.

Stigler, S.M. (1986), *The History of Statistics*. Cambridge, MA: Harvard Universit Press.

Stock, J.H., and M.W. Watson (1989), "Interpreting the Evidence on Money-Income Causality," *Journal of Econometrics* 40, 161 – 181.

Stock, J.H., and M.W. Watson (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econo-*

metrica 61, 783 – 820.

Vella, F., and M. Verbeek (1998), "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men," *Journal of Applied Econometrics* 13, 163 – 183.

Wald, A. (1940), "The Fitting of Straight Lines if Both Variables Are Subject to Error," *Annals of Mathematical Statistics* 11, 284 – 300.

Wallis, K. E. (1972), "Testing for Fourth-Order Autocorrelation in Quarterly Regression Equations," *Econometrica* 40, 617 – 636.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817 – 838.

White, H. (1984), *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

White, M. J. (1986), "Property Taxes and Firm Location: Evidence from Proposition 13," in *Studies in State and Local Public Finance*. Ed. H. S. Rosen, 83 – 112. Chicago: University of Chicago Press.

Whittington, L. A., J. Alm, and H. E. Peters (1990), "Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States," *American Economic Review* 80, 545 – 556.

Wooldridge, J. M. (1989), "A Computationally Simple Heteroskedasticity and Serial Correlation-Robust Standard Error for the Linear Regression Model." *Economics Letters* 31, 239 – 243.

Wooldridge, J. M. (1991a), "A Note on Computing R-Squared and Adjusted R-Squared for Trending and Seasonal Data," *Economics Letters* 36, 49 – 54.

Wooldridge, J. M. (1991b), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics* 47, 5 – 46.

Wooldridge, J. M. (1994a), "A Simple Specification Test for the Predictive Ability of Transformation Models," *Review of Economics and Statistics* 76. 59 – 65.

Wooldridge, J. M. (1994b), "Estimation and Inference for Dependent Processes," Chapter 45 in *Handbook of Econometrics*, Volume 4. Ed. R. F. Engle and D. L. McFadden, 2639 – 3738. Amsterdam: North-Holland.

Wooldridge, J. M. (1995), "Score Diagnostics for Linear Models Estimated by Two Stage Least Squares," in *Advances in Econometrics and Quantitative Economics*. Ed. G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66 – 87. Oxford: Blackwell.

Wooldridge, J. M. (1999), *Econometric Analysis of Cross Section and Panel Data*. Forthcoming, Cambridge, MA: MIT Press.

术语表

A

校正 R^2：多元回归分析中拟合优度的量度，在估计误差的方差时对添加的解释变量用一个自由度来调整。

对立假设：检验虚拟假设时的相对假设。

AR（1）序列相关：时间序列回归模型中的误差遵循 AR（1）模型。

渐近偏误：见非一致性。

渐近置信区间：大样本容量下近似成立的置信区间。

渐近正态性：适当正态化后样本分布收敛到标准正态分布的估计量。

渐近性质：当样本容量无限增长时适用的估计量和检验统计量性质。

渐近标准误：大样本下生效的标准误。

渐近 t 统计量：大样本下近似服从标准正态分布的 t 统计量。

渐近方差：为了获得渐近标准正态分布我们必须用以除估计量的平方值。

渐近有效：对于服从渐近正态分布的一致性估计量，有最小渐近方差的估计量。

渐近不相关：时间序列过程中，随着两个时点上的随机变量的时间间隔增加，它们之间的相关趋于零。

衰减偏误：总是朝向零的估计量偏误；因而，有衰减偏误的估计量的期望值小于参数的绝对值。

扩充迪基-富勒检验：包含作为回归元的滞后变量的单位根检验。

自相关：见序列相关。

自回归条件异方差性（ARCH）：动态异方差性模型，即给定过去信息，误差项的方差线性依赖于过去的误差的平方。

一阶自回归过程［AR（1）］：一个时间序列模型，其当前值线性依赖于最近的值加上一个无法预测的扰动。

辅助回归：用于计算检验统计量——例如异方差性和序列相关的检验统计量——或其他任何不估计主要感兴趣的模型的回归。

平均值：n 个数之和除以 n。

B

平衡综列：对所有的横截面单元均有所有年份（或时期）的数据的综列数据集。

基组：在包含虚拟解释变量的多元回归模型中，由截距代表的组。

基期：对于指数数字，例如价格或生产指数，其他所有时期均用来作为衡量标准的时期。

基期值：指定的基期的值，用以构造指数数字；通常基本值为 1 或 100。

基准组：见基组。

贝努里随机变量：取值为 0 或 1 的随机变量。

最优线性无偏估计量（BLUE）：在所有线性、无偏估计量中，有最小方差的估计量。在高斯-马尔科夫假定下，OLS 是以解释变量样本值为条件的 BLUE。

贝塔系数：见标准化系数。

偏误：估计量的期望参数值与总体参数值之差。

偏误估计量：期望或抽样平均与假设要估计的总体值有差异的估计量。

向零的偏误：描述的是估计量的期望绝对值小于总体参数的绝对值。

二值响应模型：二值因变量的模型。

二值变量：见虚拟变量。

二项分布：n 次独立的贝努里试验中成功次数的概率分布，其中每次试验成功的概率相同。

两变量回归模型：见简单线性回归模型。

BLUE：见最优线性无偏估计量。

Breusch-Godfrey 检验：渐近正确的 AR（p）序列相关检验，以 AR（1）最为流行；该检验考虑到滞后因变量和其他不是严格外生的回归元。

Breusch-Pagan 检验：将 OLS 残差的平方对模型中的解释变量做回归的异方差性检验。

C

因果效应：一个变量在其余条件不变情况下的变化对另一个变量产生的影响。

截尾回归模型：截去因变量高于或低于某已知阈值的数据的多元回归模型。

中心极限定理：这是概率理论得出的关键结论。它表明，独立的随机变量或甚至是弱相依随机变量，用标准离差进行标准化之后，随着样本容量的增大其分布趋于标准正态分布。

其余条件不变：其他所有相关因素均保持固定不变。

χ^2 分布：将独立的标准正态随机变量的平方相加所得到的概率分布。总和的项数等于分布的自由度。

邹至庄检验：检验不同组（如男性和女性）或不同时期（如政策变化之前和之后）的回归参数相等的 F 检验。

经典含误差变量（CEV）：观测的量度等于实际变量加上一个独立的或至少不相关的测量误差的测量误差模型。

经典线性模型：全套经典线性模型假定下的复线性回归模型。

经典线性模型（CLM）假定：对多元回归分析的理想假定集：对横截面分析为假定 MLR.1～MLR.6，对时间序列分析为假定 TS.1～TS.6。假定包括对参数为线性、无完全共线性、零条件均值、同方差、无序列相关和误差正态性。

聚类效应：对聚类中的所有单位（通常为人）有共同影响的未观测到的效应。

聚类样本：自然聚类或群组（通常由人组成）的样本。

科克伦-奥克特（CO）估计：估计含 AR（1）误差和严格外生解释变量的多元线性回归模型的一种方法；与普莱斯-温斯登估计不同，科克伦-奥克特估计不使用第一期的方程。

判定系数：见 R-平方。

协积：其概念为两个一阶自积的序列的线性组合是零阶自积的。

合成误差：综列数据模型中不随时间而变的未

观测到的效应与特异性误差之和。

条件分布：一个随机变量在给定其他一个或多个随机变量的值的条件下所服从的概率分布。

条件期望：一个随机变量（称为因变量或被解释变量）的期望或平均值，它由其他一个或多个随机变量（称为自变量或解释变量）的值来决定。

条件预测：假定某些解释变量的将来值为已知的预测

条件方差：一个随机变量在给定其他一个或多个随机变量条件下的方差。

置信区间（CI）：用于构造随机区间的规则，以使所有数据集中的某一百分比（由置信水平决定）给出包含总体值的区间。

置信水平：我们想要可能的样本置信区间包含总体值的百分比；95%是最常见的置信水平，90%和99%也用。

一致性估计量：当样本容量无限增长时，依概率收敛到总体参数的估计量。

一致性检验：当样本容量无限增长时，在对立假设条件下，拒绝虚拟假设的概率收敛到1的检验。

不变弹性模型：因变量关于解释变量的弹性为常数的模型；在多元回归中，两者均以对数形式出现。

同期外生回归元：在时间序列或综列数据应用中，与同期误差项不相关但对其他时期则不一定的回归元。

连续随机变量：取任何特定值的概率都为零的随机变量。

控制组：在项目评估中，不参与该项目的组。

控制变量：见解释变量。

角点解：在严格正值上大致连续的非负因变量，以某种规则取零值。

相关系数：两个随机变量之间线性相依的衡量，它与测量的单位无关，且在-1～1之间。

计数变量：取值为非负整数的变量。

协方差：两个随机变量之间线性相依的衡量。

协方差平稳：时间序列过程，其均值、方差为常数，且序列中任意两个随机变量之间的协方差仅与它们的间隔有关。

协变量：见解释变量。

临界值：在假设检验中，用于与检验统计量比较来决定是否拒绝虚拟假设的值。

横截面数据集：在给定时点上从总体中收集的数据集。

累积分布函数（cdf）：给出随机变量小于或等于某一指定实数的概率的函数。

D

数据截尾：当因为我们在上（或下）阈值处只知道结果高于（或低于）阈值，而不是总能观测到因变量的结果时，这一情况便产生了。（又见截尾回归模型。）

数据频率：收集时间序列数据的区间。年度、季度和月度是最常见的数据频率。

数据开采：用相同的数据集估计众多的模型来寻找“最好”的模型的做法。

戴维森-麦金农检验：用于检验相对于非嵌套对立假设的模型的检验；它可用相争持模型中得出的拟合值的t检验来实现。

自由度（*df*）：在多元回归模型分析中，观测值的个数减去待估参数的个数。

分母自由度：F检验中无约束模型的自由度。

因变量：在多元回归模型（和其他各种模型）中被解释的变量。

描述性统计量：用于概括一组数字的统计量；样本平均值、样本中位数、样本标准差是最常见的。

除季节因素：从月度或季度时间序列中除去季节成分。

除趋势：从时间序列中除去趋势的做法。

迪基-富勒分布：用以检验单位根虚拟假设的t统计量的极限分布。

迪基-富勒（DF）检验：在AR（1）模型中单位根虚拟假设的t检验。（又见增广迪基-富勒检验。）

斜率级差：所描述的是模型中某些斜率参数，因组或时期的不同而不同。

差异中的差分估计量：用两时期数据进行政策

分析产生的估计量。估计量的一个形式适用于独立的混合横截面，另一个适用于综列数据集。

递减边际效应：当解释变量的值增加，解释变量的边际效应越来越小。

离散随机变量：取值的个数为有限的或可数无限的随机变量。

分布滞后模型：将因变量与自变量的当前和过去值联系起来的时间序列模型。

扰动：见误差项。

向下偏误：估计量的期望值低于参数的总体值。

虚拟因变量：见二值响应模型。

虚拟变量：取值为0或1的变量。

虚拟变量回归：在综列数据背景下，对每个横截面单元有一个虚拟变量连同其余解释变量一起进行的回归。该回归得到固定效应估计量。

虚拟变量陷阱：自变量中包含了过多的虚拟变量造成的错误；当模型中既有整体截距又对每一组都设有一个虚拟变量时，该陷阱就产生了。

久期分析：截尾回归模型的一个应用，其中因变量是某一事件发生之前消逝的时间，例如失业者再就业之前的时间。

德宾-沃森（DW）统计量：在经典线性回归假设下，用于检验时间序列回归模型的误差项中的一阶序列相关的统计量。

动态完整模型：设更多的滞后因变量，或设更多的滞后解释变量都无助于解释因变量的均值的时间序列模型。

E

计量经济模型：将因变量与一组解释变量和未观测到的扰动联系起来的方程，方程中未知的总体参数决定了各解释变量在其余条件不变下的效应。

经济模型：从经济理论或不那么正规的经济原因中得出的关系。

经济显著性：见实际显著性。

弹性：给定一个变量在其余条件不变下增加1%，另一个变量的百分比变化。

经验分析：用正规计量分析中的数据检验理论、估计关系式或确定政策效应的研究。

内生性：用于描述存在内生解释变量的术语。

内生解释变量：在多元回归模型中，由于遗漏变量、测量误差或联立性的原因而与误差项相关的解释变量。

内生样本选择：非随机样本选择，其选择直接地或通过方程中的误差项与因变量相联系。

内生变量：在联立方程模型中，由系统内方程决定的变量。

恩格尔-葛兰杰两步程序：估计误差纠正模型的两步法，其中第一阶段估计协积参数，第二阶段估计误差纠正参数。

误差纠正模型：一阶差分形式并且包含误差纠正项的时间序列模型，它起到将两个I(1)序列带回到长期均衡的作用。

误差项：在简单或多元回归方程中，包含了未观测到的影响因变量的因素的变量。误差项也可能包含被观测的因变量或自变量中的测量误差。

误差方差：多元回归模型中误差项的方差。

含误差变量：因变量或某些自变量有测量误差的情况。

估计值：对一特定样本数据估计量所取的数字值。

估计量：组合数据以得出总体参数的数字值的规则；规则的形式不依赖于获得的特定样本。

事件研究：事件（例如政府规制或经济政策的变化）对结果变量的效应的计量分析。

排除一个有关变量：在多元回归分析中，遗漏了一个对因变量有非零偏效应的变量。

排斥性约束：说明某些变量被排斥在模型之外（或具有零总体参数）的约束。

外生解释变量：与误差项不相关的解释变量。

外生样本选择：或者依赖外生解释变量，或者与所感兴趣的模型中的误差项不相关的样本选择。

外生变量：任何与所感兴趣的模型中的误差项不相关的变量。

期望值：随机变量的分布中对中心趋势的衡量，随机变量包括估计量。

实验：在概率中用于表示结果不确定的事件的通用术语。在计量分析中，它表示通过将个体随机分配到对照组和处理组来收集数据的情况。

实验数据：通过进行受控制的实验获得的数据。

试验组：见处理组。

解释平方和（SSE）：多元回归模型中拟合值的总样本变异。

被解释变量：见因变量。

解释变量：在回归分析中，用于解释因变量中的变异的变量。

指数函数：一个数学函数，其定义为对所有值有递增的斜率，但变化的比例固定。

指数平滑：预测变量的一个简单方法，它用到该变量的所有以前结果的某种加权。

指数趋势：有固定增长率的趋势。

F

F **分布**：两个独立 χ^2 分布的随机变量分别除以各自的自由度后相除所获得的概率分布。

F **统计量**：在多元回归模型中，用于检验关于参数的多重假设的统计量。

可行的 GLS（FGLS）估计量：方差或相关参数未知，因而必须先进行估计的 GLS 程序。（又见广义最小二乘估计量。）

有限分布滞后（FDL）模型：允许一个或多个解释变量对因变量有滞后效应的动态模型。

一阶差分：对相邻时期做差分所构成的对时间序列的转换，即用后一时期减去前一时期。

一阶差分方程：在时间序列或综列数据模型中，因变量和自变量均取一阶差分的方程。

一阶差分估计量：在综列数据背景下，应用于不同时期数据的一阶差分的混合 OLS 估计量。

一阶条件：用于求解 OLS 估计值的一组线性方程。

拟合值：在各观测中将自变量的值插入 OLS 回归线时，所得到的因变量的估计值。

固定效应：见未观测到的效应。

固定效应变换：从综列数据得到消除时间均值后的数据。

预测误差：实际结果与预测结果之差。

预测区间：在预测中，关于时间序列变量的还未实现的将来值的置信区间。

函数形式的错误设定：当模型中有被遗漏的解释变量的函数（例如二次项），或者对一个因变量或某些自变量用了错误的函数时产生的问题。

G

高斯-马尔科夫假定：一组假定（假定 MLR.1 ~ MLR.5 或假定 TS.1 ~ TS.5），在这之下 OLS 是 BLUE。

高斯-马尔科夫定理：该定理表明，在五个高斯-马尔科夫假定下（对于横截面或时间序列模型），OLS 估计量是 BLUE（在解释变量样本值的条件下）。

广义最小二乘（GLS）估计量：通过对原始模型的变换，说明了已知结构的误差的方差（异方差性）和误差中的序列相关形式或两者兼有的估计量

几何（或考伊克）分布滞后：滞后系数以几何级数递减的无限分布滞后模型。

拟合优度度量：概括一组解释变量有多好地解释了因变量或响应变量的统计量。

葛兰杰因果性：一种有限意义的因果关系，其中一个序列（x_t）的过去值有助于预测另一个序列（y_t）在控制 y_t 的过去值之下的将来值。

增长率：时间序列中相对于前一时期的比例变化。可将它近似为对数差分或以百分比形式报导。

H

Heckit 方法：一种经济计量程序，用于纠正由于偶然截尾或以一些其他非随机形式缺失数据所导致的样本选择偏误。

异质性偏误：由于遗漏异质性（或遗漏变量）所致的 OLS 偏误。

异方差性：给定解释变量，误差项的方差不为常数。

未知形式的异方差性：以一未知的任意形式依赖于解释变量的异方差性。

异方差—稳健 F 统计量：对未知形式的异方差性而言（渐近）稳健的 F 统计量。

异方差—稳健 LM 统计量：对未知形式的异方差性而言（渐近）稳健的 LM 统计量。

异方差—稳健标准误：对未知形式的异方差性而言（渐近）稳健的标准误。

异方差—稳健 t 统计量：对未知形式的异方差性而言（渐近）稳健的 t 统计量。

高持续性过程：时间序列过程，其中遥远的将来的结果与当前的结果高度相关。

同方差性：回归模型中的误差在解释变量条件下具有不变的方差。

假设检验：虚拟假设或维持假设相对于对立假设的统计检验。

I

识别方程：对参数可进行一致性估计的方程，尤其是在含有内生解释变量的模型中。

特异性误差：在综列数据模型中，既在各单元（如个体、公司或城市）之间变化，又随时间变化的误差。

即期弹性：在分布滞后模型中，给定自变量增加1%因变量的即时的百分比变化。

即期乘数：见即期倾向。

即期倾向：在分布滞后模型中，自变量增加一个单位因变量的即时的变化。

偶然截尾：样本选择问题，指一个变量（通常为因变量）仅对于另一个变量的某些结果可以被观测到。

包含一个无关变量：用OLS估计方程时，回归模型中包含了总体参数为零的解释变量。

非一致性：估计量的概率极限与参数值之差。

独立随机变量：联合分布等于边际分布之积的随机变量。

自变量：见解释变量。

独立混合横截面：将独立随机样本在不同时点上进行混合所得到的数据集。

指数：关于经济行为（例如生产或价格）总量信息的统计量。

无限分布滞后（IDL）模型：自变量的变化对因变量产生的冲击影响到无限将来的分布滞后模型。

影响重大的观测值：见奇异值。

信息集：在预测中，可观测其以往数据来形成预测的变量集。

样本内准则：选取预测模型的一种标准，用以获得参数估计值的样本中的拟合优度为基础。

工具变量（IV）：在含有内生解释变量的方程中，不出现在方程中、与方程中的误差无关且与内生解释变量（偏）相关的变量。

工具变量（IV）估计量：当一个或多个内生解释变量可获得工具变量时，用于线性模型中的估计量。

一阶自积［I(1)］：需要做一阶差分来得到I(0)过程的时间序列过程。

零阶自积［I(0)］：平稳、弱独立时间序列过程，当用于回归分析时，它满足大数定律和中心极限定理。

交互作用：回归模型中为两个解释变量的乘积的自变量。

截距参数：复线性回归模型中，给出当所有自变量都为零时因变量的期望值的参数。

截距的变动：回归模型中的截距，因组或时期的不同而不同。

因特网：全球计算机网络，可用于获得信息和下载数据库。

区间估计量：用数据得到总体参数上下界限的规则。

逆米尔斯比率：添加到多元回归模型中以消除样本选择偏误的一项。

J

联合分布：决定了涉及两个或多个随机变量的概率结果的概率分布。

联合假设检验：一个模型中包含不止一个对参数的约束的检验。

联合统计显著性：两个或多个解释变量具有零总体系数的虚拟假设以一个选定的显著性水平被拒绝

恰好识别方程：对有内生解释变量的那些模型来说，一个能被识别，但若减少一个工具变量则不能被识别的方程。

L

滞后分布：在无限或有限分布滞后模型中，把滞后系数表示为滞后长度的函数。

滞后因变量：等于以前时期的因变量的解释变量

滞后内生变量：在联立方程模型中，某一内生变量的滞后值。

拉格朗日乘数统计量：仅在大样本下为正确的检验统计量，它可用于在不同的模型设定问题中检验遗漏变量、异方差性和序列相关。

大样本性质：见渐近性质。

潜在变量模型：这样一个模型，其中观测到的因变量假定为潜在的或未观测到的变量的函数。

重叠期望律：概率论的一个结论，它把无条件期望和条件期望相联系。

大数（LLN）定律：说明随机样本的平均值依概率收敛到总体平均值的定理；LLN 对平稳的和弱相依的时间序列也成立。

先导与滞后估计量：I(1) 变量回归中的协积参数的估计量，它包含了解释变量的现在的、一部分过去的、一部分将来的一阶差分作为回归元。

水平值—水平值模型：因变量与自变量均为标准（或原始）形式的回归模型。

水平值—对数模型：因变量为标准形式、自变量（至少是其中一部分）为对数形式的回归模型。

似然比统计量：用最大似然估计受约束的与无约束的模型之后，可用于检验单个或多重检验的统计量。该统计量是无约束与受约束的对数似然函数之差的 2 倍。

限值因变量：其变化范围在重要的程度上受到限制的因变量或响应变量。

线性函数：自变量变化一个单位，因变量的变化为常数的函数。

线性概率模型（LPM）：响应概率对参数为线性的二值响应模型。

线性时间趋势：为时间的线性函数的趋势。

线性无偏估计量：在多元回归分析中，是因变量值的一个线性函数的那些无偏估计量。

对数函数：定义于正值自变量的一种有正的、递减斜率的数学函数。

对数—水平值模型：因变量以对数形式出现，而自变量是水平（或原始）形式的一种回归模型。

对数—对数模型：因变量和（至少一部分）解释变量都是以对数形式出现的回归模型。

对数单位模型：关于二值响应的一种模型，其中的响应概率为逻辑斯蒂函数在解释变量的线性函数上的取值。

对数—似然函数：等于对数—似然值的总和，而每个观测的对数—似然值是给定解释变量时因变量的概率密度的对数。它是被估计参数的一个函数。

长期弹性：因变量和自变量都是对数形式出现的分布滞后模型中的长期倾向。即长期弹性是在给定解释变量增长了 1% 时，被解释变量最终变化的百分比。

长期乘数：见长期倾向。

长期倾向：在一个分布滞后模型中，给定自变量的一个永久性的、一个单位的增长，因变量最终的变化量。

纵横数据：见综列数据。

损失函数：一种表示预测值与实际结果不相同所造成的损失的函数，最常见的例子是绝对值损失和平方损失。

M

边际效应：自变量有很小数量的一个改变，所造成的对因变量的影响。

鞅：给定序列的所有过去的结果时，序列的期望值等于最近一期的值的一种时间序列过程。

鞅差分序列：鞅的一阶差分。给定序列的过去值，它是不可预报的（或有零均值）。

配对样本：每个观测值都与另一个观测值相匹配的一种样本，如由丈夫和妻子或一对兄妹组成的样本。

矩阵：数据的一种排列方式。

矩阵表示法：建立在矩阵代数基础上的一种便捷的数学表示方法，用来表示和计算多元回归模型

最大似然估计（MLE）：一种应用广泛的估计方法，通过最大化对数似然函数来选择参数估计值。

均值：见期望值。

平均绝对误差（MAE）：表示预测效果的一种指标，通过计算预测误差绝对值的平均值而得到。

均方误：估计量与总体值之差的平方的期望，它等于方差加上某一偏误的平方。

测量误差：观测到的变量与多元回归方程中的变量之间的差。

中位数：在概率分布中，它是这样一个数，在它之下有50%的概率，在它之上也有50%的概率。在一个数据样本中，它是经过排序后位置居中的值。

矩估计量方法：使用与总体矩相似的样本矩计算得到的一种估计量，最小二乘法和两阶段最小二乘法都属于矩估计量方法。

微数缺测性：由阿瑟·戈德伯格（Arthur Goldberger）首先提出的一个概念，用以描述容量样本较小时计量经济学估计量的性质。

最小方差无偏估计量：在所有的无偏估计量中方差最小的那个估计量。

数据缺失：当我们没有观测到样本中某些观测（个人、城市、时期等）所对应的一些变量值的时，发生的一类数据问题。

一阶移动平均过程［MA（1）］：是由某个随机过程的当期值与一期滞后的线性函数所产生的一种时间序列过程。这个随机过程是零均值、固定方差和不相关的。

多重共线性：指多元回归模型中自变量之间的相关性；当某些相关性“很大”时，就会发生多重共线性，但对实际的大小尺度并没有明确的规定。

多重假设检验：涉及参数的多个约束条件的虚拟假设检验。

多元线性回归（MLR）模型：对参数是线性的一类模型，其中的因变量是自变量的函数加上一个误差项。

多元回归分析：在多元线性回归模型中进行估计和推断的一类分析。

多重约束：计量经济学模型中对参数的多于一个的约束条件。

超前多步预测（值）：对距现在一个时期以上的将来的一种时间序列预测（值）。

乘数测量误差：观测到的变量等于实际的观测不到的变量与一个正的测量误差的乘积时出现的一种测量误差。

N

n-R-平方统计量：参见拉格朗日乘数统计量。

自然实验：通常由政策或制度的变化无意之中所造成的经济环境的外生性变化，这种情况被称为自然实验。其中的经济环境有时由一个解释变量来表示。

自然对数：见对数函数。

名义变量：用名义或当前美元数表示的变量。

非实验数据：不是通过人为控制下的实验得到的数据。

非线性函数：斜率不是常数的函数。

非嵌套模型：没有一个模型可以通过对参数施加限制条件而被表示成另一个模型的特例的两个（或更多）模型。

非随机样本选择：不是从我们感兴趣的总体中随机抽样的一种样本选择方法。

非平稳过程：联合分布在不同的时期不是恒定不变的一种时间序列过程。

正态分布：在统计学和计量经济学中构造总体模型经常用到的一种概率分布。它的概率分布函数呈钟形。

正态性假定：经典线性模型假定之一。它是指

以解释变量为条件的误差（或因变量）有正态分布。

虚拟假设：在经典假设检验中，我们把这个假设当做真的，要求数据能够提供足够的证据才能否定它。

分子自由度：在 *F* 检验中，所检验的约束条件的个数。

O

可观测数据：见非实验数据。

OLS：见普通最小二乘法。

OLS 截距估计值：OLS 回归线的截距。

OLS 回归线：表示了因变量的预报值与自变量的值之间关系的方程，它的参数是用 OLS 估计出来的。

OLS 斜率估计值：OLS 回归线的斜率。

遗漏变量偏误：回归中遗漏了有关变量而产生的 OLS 估计量的偏误。

遗漏的变量：我们打算控制的但在估计回归模型的过程中被漏掉的那些变量。

单侧对立假设：被表述为参数大于（或小于）虚拟条件下的假设值的一种对立假设。

超前一步预测（值）：对未来下一个时期所做的预测（得到的预测值）。

单尾检验：与单侧对立假设相对的假设检验。

网上数据库：可以通过计算机互联网进入的数据库。

网上搜索服务：帮助人们根据主题、名称、标题或关键词在互联网或网上数据库中搜索资料的计算机软件。

阶条件：在有一个或多个内生解释变量的方程中识别参数的必要条件：所有外生变量的个数必须不少于解释变量的个数。

序数变量：通过排列顺序传达信息的一种数据，它们的大小本身并不说明任何问题。

普通最小二乘法（OLS）：用来估计多元线性回归模型中的参数的一种方法。最小二乘估计值通过最小化残差的平方和得到。

异常数据：在数据集中，与大量其他数据有明显区别的观测值，这种现象可能是由于误差造成的，也可能是因为它们是由与多数其他数据不同的模型产生而造成的。

样本外准则：利用在计算参数估计值的过程中未使用的那部分样本来选择预测模型的一种准则。

整体显著性：对多元回归方程中所有的解释变量所做的一种联合显著性检验。

过度分散：在对计数变量建立模型时，方差大于均值。

过度识别方程：在有内生解释变量的一组方程中，工具变量的个数大于内生解释变量个数的那个方程。

过度识别约束：在线性模型中，工具变量的个数多于内生解释变量的个数，所要求的额外的矩阵条件。

模型的过度识别：见含有一个无关变量。

P

***p* 值**：指能够拒绝虚拟假设的最低显著性水平。等价地，它也指虚拟假设不被拒绝的最大显著性水平。

综列数据：在不同时期，横截面的不断反复得到的数据集。在平衡的综列中，同样的单位在每个时期都出现。在不平衡的综列中，有些单位往往由于衰减现象而不会在每个时期都出现。

配对不相关随机变量：指每对变量都是不相关的、由两个或更多变量组成的一组变量。

参数：一种描述了总体关系的未知数值。

简约模型：为了表现任何所求特征而使用的参数个数尽可能少的模型。

偏效应：回归模型中的其他因素保持不变时，某个解释变量对因变量的影响。

正确预测的百分比：在二元响应模型中，预报的 0 或 1 与实际结果相一致的次数所占总次数的百分比。

百分数变化：等于变量变化的比例再乘以 100。

百分点变化：用百分之一作为单位度量的变量的变化。

完全共线性：在多元回归中，一个自变量是一个或多个其他自变量的线性函数。

变量缺失问题的插入解：在OLS回归中，用一个代理变量代替观测不到的缺失变量。

点预测：一个将来结果的预测值。

泊松分布：计数变量的一种概率分布。

泊松回归模型：计数因变量的一种模型。其中，以解释变量为条件的因变量，名义上被假定遵循泊松分布。

政策分析：用计量经济学模型来评估某项政策的效果的一种实证分析。

混合横截面：通常在不同时点收集到的相互独立的横截面组合而成的一个单独的数据集。

混合OLS估计：使用相互独立的混合横截面、综列数据或聚集样本的OLS估计。其中，不同时期（或组群）及不同的横截面单位的观测值被混合在一起。

总体：作为统计或计量经济分析对象的一个明确定义的组群（人、公司、城市等）。

总体模型：一种描述了总体特征的模型，特别是多元线性回归模型。

总体*R*-平方：总体中，由解释变量解释了的那部分因变量的变异。

总体回归函数：见条件期望。

检验的功效：当虚拟假设错误时拒绝它的可能性的大小；这种功效取决于对立假设下的总体参数值。

实际显著性：相对于统计显著性而言的、某个估计值的实际的或经济的重要性，用它的符号和大小来衡量。

普莱斯-温斯登（PW）估计：一种用来估计有AR（1）误差和严格外生解释变量的多元线性回归模型的方法；不同于科克伦-奥克特方法，它在估计中要用到第一个时期的方程。

前定变量：在联立方程模型中的滞后的内生变量或滞后的外生变量。

被预测变量：见因变量。

预报：把特定的解释变量的值代入所估计的模型，通常是多元回归模型中，以得到结果的一个估计值。

预测误差：实际结果与所预报的结果之间的差。

预测区间：多元回归模型中，某个因变量的未知结果的一个置信区间。

预测变量：见解释变量。

概率密度函数（pdf）：是这样一种函数：对于离散随机变量，它给出了这个随机变量落在每个值上的概率；对于连续随机变量，pdf之下的面积则给出了各种事件的概率。

概率极限：当样本规模无限增大时，估计量所趋近的值。

概率单位模型：一种二值响应模型，其中的响应概率为标准正态的累积分布函数在解释变量的一个线性函数处取值。

项目评估：用计量经济学方法求出某个私人或公共项目的不确定影响的一种评估方法。

比例变化：一个变量相对于它的初始值的变化；从数学上来讲，就是用变化量除以初始值。

代理变量：多元回归分析中，一个与观测不到的解释变量有关系但又不相同的可观测变量。

Q

二次函数：包含一个或多个解释变量的平方的函数；它反映了解释变量对因变量的逐渐变弱或增强的影响。

定性变量：描述一个个人、企业及城市等的非定量特征的变量。

拟除均值数据：在综列数据的随机影响估计中，每个时期的原始数据减去时间均值后得到的数据；对每个横截面观测都要进行这样的计算。

拟差分数据：在估计有AR（1）的序列相关的回归模型时，当期数据与前一期数据乘以AR（1）模型的参数后得到的数据之间的差。

拟实验：见自然实验。

拟似然比统计量：就像泊松回归模型中那样，经过修正的似然比统计量，用来解释可能的分布错误设定。

拟最大似然估计法：对数—似然函数可能与因

变量的实际条件分布不一致的一种最大似然估计方法。

R

$\bar{R}$-平方：见校正的 R-平方。

R-平方：在多元回归模型中，由自变量解释了的那部分因变量的样本方差之和。

R-平方形式的 F 统计量：用受约束和不受约束的模型中得到的由 R-平方表示的、用于检验排除约束条件的 F 统计量。

随机效应估计量：不可观测效应的模型中的一种可行的 GLS 估计量。其中，不可观测的效应被假定为与每个时期的解释变量都不相关。

随机效应模型：不可观测效应的综列数据模型。其中，假定不可观测的效应与每个时期的解释变量都不相关。

随机抽样：在总体中随机抽取观测值的一种抽样方法。各个单位被抽取的可能性是相同的，而且每次抽样都与其他次相互独立。

随机变量：结果不确定的一种变量。

随机游走：在这样一种时间序列中，下个时期的值等于本期值加上一个独立的（或至少是不相关的）误差项。

有漂移的随机游走：每个时期都加进一个常数（或漂移）的随机游走。

秩条件：识别有一个或多个内生解释变量的模型的充分条件。

有理分布滞后（RDL）模型：一种无限分布滞后模型，其中的滞后分布取决于相对较少的参数。

实际变量：用基期货币价值表示的变量。

诱导（简约）型方程：一种特殊的线性方程，其中的内生变量是外生变量与观测不到的误差的函数。

诱导型误差：诱导型方程中出现的误差。

回归子：见因变量。

回归误差设定检验（RESET）：在多元回归模型中，检验函数形式的一般性方法。它是一种由最初的 OLS 估计得出的拟合值的平方、三次方以及可能更高次幂的联合显著性 F 检验。

过原点回归：截距被设为零的回归分析，它的斜率通过最小化残差的平方和求出。

回归元：见解释变量。

拒绝区域：使得虚拟假设被拒绝的一组检验统计量的值。

拒绝法则：在假设检验中，决定在什么情况下拒绝虚拟假设并支持对立假设的法则。

残差：实际值与拟合（或预报）值之间的差；样本中的每次观测都有一个相应的残差，它们被用来计算 OLS 回归线。

残差分析：在估计多元回归模型后，对某次特定观测的残差的符号和大小所作的研究。

残差平方和：见残差的平方和。

响应概率：在二值响应模型中，以解释变量为条件的因变量取值为 1 的概率。

响应变量：见因变量。

受约束的模型：在假设检验中，施加所有虚拟假设所要求的约束条件后得到的模型。

均方根误（RMSE）：多元回归分析中回归标准误的另一个名称（仅当期望值等于实测值——译者注）。

S

样本均值：n 个数之和除以 n，它度量了中心化倾向。

样本相关性：关于两个随机变量的结果，样本协方差除以两个样本标准差的乘积就是样本相关性。

样本协方差：两个随机变量之间的总体协方差的一个无偏估计量。

样本回归函数：见 OLS 回归线。

样本选择偏误：由于使用了内生样本选择方法得到数据而产生的 OLS 估计量的偏误。

样本标准差：总体标准差的一个一致估计量。

样本方差：总体方差的一个无偏的、一致估计量。

抽样分布：一个估计量相对于所有可能的抽样结果的概率分布。

抽样方差：一个估计量的抽样分布的方差，它度量了抽样分布的分散程度。

得分统计量：见拉格朗日乘数统计量。

季节性虚拟变量：一组用来表示季节或月份的虚拟变量。

季节性：月度或季度时间序列具有的均值随着一年中季节的不同而系统性变化的特点。

季节性调整：用某种统计程序，可能是对季节性虚拟变量做回归，来消除月度或季度时间序列中的季节性成分。

选择性样本：不是通过随机抽样得到的，而是依据某些观测得到或观测不到的特征选取的样本数据。

半弹性：自变量的一个单位的增长导致的因变量的变化的百分比。

敏感性分析：检验关键解释变量的估计效应和统计显著性是否对添加其他解释变量、函数形式、去掉潜在的异常观测和不同的估计方法敏感的过程。

序列相关：在时间序列或综列数据模型中，不同时期的误差之间的相关性。

序列相关—稳健标准误：不管模型中的误差是否序列相关，都（渐近）生效的估计量的标准误。

序列不相关：在时间序列或综列数据模型中，不同时间的误差两两之间不相关。

短期弹性：因变量和自变量都以对数形式出现的分布滞后模型中的即期倾向。

显著性水平：假设检验中发生第Ⅰ类错误的概率。

简单线性回归模型：因变量只是一个自变量和一个误差项的线性函数的模型。

联立性：在多元线性回归模型中，至少有一个解释变量与因变量一同决定。

联立性偏误：用OLS估计联立方程模型中的一个方程时出现的偏误。

联立方程模型（SEM）：两个或多个内生变量同时被决定的一种模型，其中的每个内生变量都可能是其他内生变量和外生变量以及误差项的函数。

斜率参数：多元回归模型中的自变量的系数。

Spreadsheet：用来输入和加工数据的计算机软件。

谬误相关：不是因为二者有因果关系，可能是因为它们都受另一个观测不到的因素影响，所导致的两个变量之间的相关性。

谬误回归问题：如果回归分析表明两个或多个无关时间序列具有一定关系，而其原因仅仅因为它们每个都有趋势，或都是自积时间序列（如随机游走），或上面两种情况同时出现，这种问题就是谬误回归问题。

稳定的AR（1）过程：滞后变量的系数绝对值小于1时的AR（1）过程。序列中的两个随机变量的相关性，随着它们之间的时间间隔不断增大，以几何级数趋近于零。

标准差：衡量随机变量分布的分散程度的常用指标。

$\hat{\beta}_j$ 的标准差：衡量 $\hat{\beta}_j$ 抽样分布的分散程度的常用指标。

$\hat{\beta}_j$ 的标准误：$\hat{\beta}_j$ 抽样分布的标准差的估计值。

估计值的标准误：见回归的标准误。

回归的标准误（SER）：多元回归分析中的总体误差的标准差的估计值，等于残差平方和的平方根除以自由度。

标准正态分布：均值为0、方差为1的正态分布。

标准化系数：一种回归系数，它度量了自变量增加一个标准差时，因变量的改变是其标准差的倍数。

标准化的随机变量：一个随机变量减去它的期望值，然后再除以它的标准差后得到的新的变量。新变量的均值为0，方差为1。

静态模型：只有当期的解释变量影响因变量的一种时间序列模型。

平稳过程：边际和所有的联合分布都不随时间变化的一种时间序列过程。

统计推断：对总体参数进行假设检验的行为。

统计上异于零：见统计上显著。

统计上不显著：在选定的显著性水平上，无法拒绝总体参数等于零的虚拟假设。

统计上显著：在选定的显著性水平上，相对于特定的对立假设，拒绝总体参数等于零的虚拟

假设。

随机过程：标注了时间的一系列随机变量。

严格外生性：时间序列或综列数据模型中的假定之一，在解释变量是严格外生的情况下成立。

严格外生的：时间序列或综列数据模型中的解释变量的一个特点：以所有时期的解释变量为条件的、任何时期的误差项都具有零均值。更宽松的一种说法是用相关性为零来表述的。

强相依：见高度持续过程。

结构方程：由经济理论或不那么正式的经济推理推导出的方程。

结构误差：结构方程中的误差项，该结构方程可以是联立方程模型中的一个方程。

结构参数：结构方程中出现的参数。

残差平方和：多元回归模型中所观测的 OLS 残差的平方和。

求和运算符：用∑表示的一个符号，用来表示对一组数据的求和运算。

T

***t* 分布**：一个标准正态随机变量与一个独立的 χ^2 方随机变量的平方根之比的分布，这个 χ^2 随机变量首先要除以它的自由度 df。

***t* 比率**：见 t 统计量。

***t* 统计量**：用来对计量经济学模型中关于参数的单个假设进行检验的一种统计量。

检验统计量：对每个样本结果都产生一个数字的假设进行检验的原则。

文本编辑软件：用来编辑文本文件的计算机软件。

文本（ASCII）文件：一种通用的文件格式，它可以在数字化的计算机平台之间传输。

除时间均值数据：从每个时期的数据中减去横截面单元在时间上的均值后得到的综列数据。

时间序列数据：搜集到的一个或多个变量在不同时间上的数据。

时间序列过程：见随机过程。

时间趋势：时间的函数，它是趋势时间序列过程的期望值。

Tobit 模型：因变量取零的概率是正的，并大致连续地分布在正数集上的一种模型。

顶端编码：一种删改数据的方法。如果一个变量高于给定的界限，它的值就不报告出来，我们只知道它至少和界限值一样大。

总平方和（SST）：因变量相对于它的样本均值的总样本变异。

处理组：在项目评估中，参与这一项目的群体。（也见实验群组。）

趋势过程：期望值是时间的增函数或减函数的时间序列过程。

趋势-平稳过程：在除掉了时间趋势后变得平稳的过程。毫无疑问，除掉了趋势的序列是弱相依的。

断尾回归模型：横截面数据的一种经典线性回归模型。在这种模型中，以知道的因变量的结果为基础，采用除掉一部分总体的抽样方法。

真实模型：表示因变量与有关自变量及一个干扰项之间关系的真实的总体模型。在这个模型中，零条件均值假定成立。

两阶段最小二乘（2SLS）估计量：做内生解释变量对所有外生变量的回归，得到的拟合值就是内生解释变量的 IV，这种方法计算出来的估计量就是两阶段最小二乘估计量。

双侧对立假设：总体参数既可以大于又可以小于虚拟假设提出的值的一种检验方法。

双尾检验：相对于双侧对立检验的检验方法。

第Ⅰ类错误：在虚拟假设正确时，拒绝了虚拟假设。

第Ⅱ类错误：在虚拟假设错误时，却没有拒绝它。

U

不平衡综列：某些年度（或时期）的数据从一些横截面单元中丢失了的综列数据。

无偏估计量：期望值（或抽样分布的均值）等于总体值（与总体值的大小无关）的估计量。

非条件预测（值）：不以知道也不以假定未来解释变量的值为条件的预测（值）。

不相关随机变量：相互之间没有线性关系的随

机变量。

设定不足的模型：见忽略一个有关的变量。

不可识别的方程：有一个或多个内生解释变量，且找不到足够多的工具变量来识别其相应参数的方程。

单位根过程：当期值等于前一个时期的值加上一个弱相依的干扰项的一种高度持续的时间序列过程。

观测不到的影响：在综列数据模型中，误差项中不随时间而变化的、未被观测到的变量。对于聚类样本来说，它就是这个聚类中所有的单元共有的一个未被观测到的变量。

观测不到的影响模型：一种使用综列数据或聚类样本的模型，它的误差项包含观测不到的影响。

观测不到的异质性：见观测不到的影响。

无约束模型：在假设检验中，对参数没有任何限制条件的模型。

向上偏误：估计量的期望值大于总体参数的值。

V

方差：表示随机变量分布的分散程度的一项指标。

预测误差的方差：当以估计的多元回归方程为基础来预报因变量的一个将来值时，产生的误差的方差。

向量自回归（VAR）模型：关于两个或多个时间序列的一种模型，其中每个变量都被表示成所有变量的过去值的线性函数，再加上一个干扰项；在给定可观测变量的所有过去值时，这个干扰项的均值为零。

W

弱相依：在时间序列过程中，表示随机变量在不同时期的两个值之间的相互依赖性质的指标（比如相关性），如果这一依赖性随着时间间隔的增大而减小，这个时间序列就是弱相依的。

加权最小二乘（WLS）估计量：用来校正某种已知形式的异方差的估计量。其中，每个残差的平方都得到一个等于误差的（估计的）方差的倒数的权重。

怀特检验：异方差的一种检验方法，涉及做OLS残差的平方对OLS拟合值和拟合值的平方的回归。这种检验方法的最一般的形式是，做OLS残差的平方对解释变量、解释变量的平方和所有非多余的解释变量间的交叉乘积的回归。

组内估计量：见固定效应估计量。

组内变换：见固定效应变换。

Y

年虚拟变量：对于有时间序列成分的数据集，在相关年度等于1而在所有其他年度都等于零的虚拟（二元）变量。

Z

零条件均值假定：多元回归分析中很关键的一个假定。它的含义是，给定解释变量的所有值时，误差的期望值都等于零。（见假定MLR.3，TS.2和TS.2′。）

索引

Index	索引
A	
ability, using two test scores as indicators of, Example 15.6	例 15.6,用两个检验得分作为能力的标示,482
adjusted R-squared	调整 R-平方,192
to choose between nonnested models	在非嵌套模型之间的选择,193
air pollution, housing prices and, Example 4.5	例 4.5,住房价格和空气污染,128
algebraic properties of OLS statistics	OLS统计量的代数性质,37
alternative hypothesis	对立假设,118,724
annual U.S. inflation, unit root test for, Example 18.3	例18.3,美国的年度通货膨胀的单位根检验,582
antidumping filings	反倾销调查
and chemical imports, Example 10.5	例 10.5,~与化学品进口,329
effects of, Example 10.11	例 10.11,~的影响,341
AR(1)	AR(1)
AR(1) errors	AR(1) 误差
feasible GLS estimation with	对~的可行 GLS 估计,389

2SLS with	含 AR(1) 误差的 2SLS,488
AR(1) model	AR(1) 模型
Example 11.3	例 11.3,354
feasible GLS estimation of	～的可行 GLS 估计,389
obtaining the best linear unbiased estimator in	在～中得到最优线性无偏估计量,387
AR(1) serial correlation	AR(1) 序列相关
after 2SLS, testing for	在 2SLS 后对～的检验,487
in the minimum wage equation, testing for, Example 12.2	例 12.2,在最低工资方程中对～的检验,384
in the Phillips curve, Example12.1, testing for,	例 12.1,在菲利普斯曲线中对～的检验,381
with strictly exogenous regressors, t test for,	含有严格外生回归元时对～的 t 检验,380
with strictly exogenous regressors, testing for	对含有严格外生回归元～的检验,381
without strictly exogenous regressors, testing for	对不含严格外生回归元～的检验,383
feasible GLS with heteroskedasticity and	含有异方差性和～的可行 GLS,403
AR(q) serial correlation, testing for	对 AR(q) 序列相关的检验,385
AR(3) serial correlation, testing for, Example 12.3	例 12.3,对 AR(3) 序列相关的检验, 385
ARCH	ARCH,400
in stock returns, Example 12.9	例 12.9,股票收益的～,402
arrest records, explaining, Example 3.5	例 3.5,对拘捕次数的解释 ,80
assumptions	假定
for fixed and random effects(Assumption FE.1 - FE.7) (Assumption RE.3 - RE.5)	对确定和随机效应的～(假定 FE.1 - FE.7)(假定 RE.3 RE.5),459 - 460
for pooled OLS using first differences (Assumption FD.1 - FD.7)	对用一阶差分的混合 OLS 的～(假定 FD.1 - FD.7),439 - 440
zero conditional mean	零条件均值～,25
asymptotic bias	渐近偏误,165
asymptotic confidence interval	渐近置信区间,170
for nonnormal populations	非正态性总体的～,772
asymptotic critical values for cointegration test	协积检验的渐近临界值
linear time trend, Table 18.5	表 18.5,线性时间趋势的～,589
no time trend, Table 18.4	表 18.4,无时间趋势的～,588
asymptotic critical values for unit root t test	单位根 t 检验的渐近临界值
linear time trend, Table 18.3	表 18.3,线性时间趋势的～,583
no time trend, Table 18.2	表 18.2,无时间趋势的～,580
asymptotic efficiency of OLS	OLS 的渐近有效性,173
Theorem 5.3	～定理 5.3, 175
asymptotic normality	渐近正态性,168,711,712
and large sample inference	～和大样本推断,167

of OLS, Theorem 11.2	定理 11.2,OLS 的～, 355
of OLS, Theorem 5.2	定理 5.2,OLS 的～,168
asymptotic or large sample properties of estimators	估计量的渐近或大样本性质,708
asymptotic properties	渐近性质,162
of OLS	OLS 的～,352
asymptotic standard error	渐近标准误,170
in limited dependent variable models	限值因变量模型中的～,569
asymptotic t statistics	渐近 t 统计量,170
asymptotic tests for nonnormal populations	非正态总体的渐近检验,729
asymptotic variance	渐近方差,168
asymptotically efficient	渐近有效,173
asymptotically uncorrelated	渐近不相关,349
attendance on final exam performance, effects of, Example 6.3	例 6.3,出勤率对期末成绩的影响,190
attenuation bias	衰减偏误,295
augmented Dickey-Fuller test	增广迪基-富勒检验,581
autocorrelation	自相关,321
autoregressive conditional heteroskedasticity (ARCH)	自回归条件异方差(ARCH),400
autoregressive process of order one [AR(1)]	一阶自回归过程[AR(1)],350
auxiliary regression	辅助回归,172
average	平均,644

B

base group	基组,213
bade period	基期,327
base value	基值,327
basic definitions	基本定义,744
benchmark group	基准组,213
Bernoulli (or binary) random variables	贝努里(或二值)随机变量,665
best linear unbiased estimator (BLUE)	最优线性无偏估计量(BLUE),101
beta coefficients	贝塔系数,181,182
bias	偏误,703
of an estimator	一个估计量的～,703
summary of, Table 3.2	表 3.2,对～的总结,90
biased towards zero	向零的偏误,91
binary dependent variable: the linear probability model	二值因变量:线性概率模型,232
binary response model	二值响应模型,233,530
binary variable	二值变量,211

binomial distribution	二项式分布,671
birth weight equation	出生体重方程
parents' education in, Example 4.9	例 4.9,～中父母的受教育程度,146
standard errors in, Example 5.2	例 5.2,～的标准误,170
birth weight, estimating the effect of smoking on, Example 15.3	例 15.3,估计吸烟对出生体重的影响,471
bivariate linear regression model	双变量线性回归模型,22
Breusch-Godfrey test	布罗施-戈弗雷检验,386
Breusch-Pagan test for heteroskedasticity (BP test)	对异方差性的布罗施-培甘检验(BP 检验),257
business investments, effect of enterprise zones on	企业区对企业投资的影响,727
C	
campaign expenditures, voting outcomes and	选举结果与竞选支出
Example 2.5	例 2.5, 34
Example 2.9	例 2.9, 41
campus crime and enrollment, Example 4.4	例 4.4,校园犯罪和注册人数,126
causal effect	因果效应,13
causality and the notion of ceteris paribus in econometric analysis	计量经济分析中的因果性和其他条件不变的概念,13
censored and truncated regression models	截取和断尾回归模型,551
censored regression estimation of criminal recidivism, Table 17.4	表 17.4,对罪犯累犯的截取回归估计,554
censored regression model	截取回归模型,551,552
central limit theorem(CLT)	中心极限定理(CLT) ,712
central tendency	中心趋势
expected value, measure of	对～的期望值度量,672
the median, another measure of	对～的另一种中位数度量,676
CEO compensation and firm performance, Example 6.4	例 6.4,CEO 的酬金和企业的业绩,195
CEO salaries, predicting	预测 CEO 薪水
Example 6.7	例 6.7,203
Example 6.8	例 6.8,204
CEO salary and firm sales, Example 2.11	例 2.11,CEO 的薪水与企业的销售额,44
CEO salary and return on equity	CEO 的薪水与股票收益
Example 2.3	例 2.3,32
Example 2.6	例 2.6,36
Example 2.8	例 2.8,40
ceteris paribus	其他条件不变情况下,13,73,647
chemical imports, antidumping filings and, Ex-	例 10.5,反倾销与化工进口,329

ample 10.5	
chi-square distribution	χ平方分布,693
Figure B.9	图B.9,694
properties of	～的性质,753
Chow statistic	邹至庄统计量,231
Chow test for structural change across time	跨时期结构变动的邹至庄检验,413
city crime levels, effect of law enforcement on, Example 1.5	例1.5,执法对城市犯罪水平的影响,16
city crime rates, Example 9.4	例9.4,城市犯罪率,289
city unemployment rates, Example C.1	例C.1,城市失业率,701－702
classical errors-in-variables(CEV)	经典的变量中误差(CEV),295
classical hypothesis testing	经典假设检验,131
classical linear model	经典线性模型,114
classical linear model (CLM) assumptions	经典线性模型(CLM)假定,114
under	在～下,322
cluster effect	群集效应,454
Cochrane-Orcutt (CO) estimation	科克伦-奥克特(CO)估计,389
in the event study, Example 12.4	例12.4,事件研究中的～,390
coefficient of determination	决定因素的系数,39
coefficients on dummy explanatory variables when the dependent variable is $\log(y)$, interpreting	当因变量是$\log(y)$时对虚拟解释变量的系数的解释,218
cointegrating parameter for interest rates	利率的协积参数,591
cointegration	协积,586
between fertility and personal exemption, Example 18.5	例18.5,生育率与个人税收豁免之间的～,589
cointegration test	协积检验
linear time trend, asymptotic critical values for, Table 18.5	表18.5,线性时间趋势下～的渐近临界值,589
no time trend, asymptotic critical values for, Table 18.4	表18.4,无时间趋势时～的渐近临界值,588
college GPA	大学GPA
confidence interval for future, Example 6.6	例6.6,将来～的置信区间,201
confidence interval for predicted, Example 6.5	例6.5,预测～的置信区间,199
determinants of, Example 3.1	～的决定因素,73
Example 3.4	例3.4,80
Example 4.3	例4.3,125
effects of computer ownership on, Example 7.2	例7.2,拥有计算机对～的影响,216
college proximity as an IV for education, using, Example 15.4	例15.4,利用大学近似作为受教育水平的工具,474
collinearity, no perfect	不完全的共线性

Assumption TS.3	假定 TS.3,319
Assumption TS.3′	假定 TS.3′,352
column vector	列向量,745
commute time, effect of freeway width on, Example C.9	例 C.9,通道宽度对上下班时间的影响,735
compact disks, demand for, Example A.2	例 A.2,对密纹唱碟的需求,647
components of the OLS variances: multicollinearity	OLS 方差的组成:多重共线性,94
composite error	合成误差,421
term	～项,449
conceptual (or theoretical) framework	概念(或理论)框架,627
conditional distribution	条件分布,669,671
features of joint and	联合和～的性质,680
conditional expectation	条件期望,684
property CE.1,CE.2, and CE.3	性质 CE.1,CE.2 和 CE.686
property CE.4,CE.4′, and CE.5	性质 CE.4,CE.4′和 CE.5.687
property CE.6	性质 CE.6,688
conditional forecast	条件预测,595
conditional variance	条件方差,688
property CV.1	性质 CV.1,688
confidence interval (CI)	置信区间(CI) ,134,715
and hypothesis testing, relationship between	～与假设检验之间的关系,734
for future college GPA, Example 6.6	例 6.6,将来大学 GPA 的～,201
for predicted college GPA, Example 6.5	例 6.5,预测大学 GPA 的～,199
for prediction	预测的～,198
for the mean, from a normally distributed population	一个正态分布总体的均值的～,718
simple rule of thumb for a 95%	构造一个 95% 的～的简单经验法则,772
consistency	一致性,162,163,708,709
unbiasedness and	无偏性和～,376
consistency of OLS	OLS 的一致性
Theorem 11.1	定理 11.1,353
Theorem 5.1	定理 5.1,164
consistent estimator	一致估计量,709
sampling distributions of, Figure C.3	图 C.3,～的抽样分布,709
consistent test	一致检验,736
constant elasticity demand function, Example A.5	例 A.5,常弹性需求函数,655
constant elasticity model	常弹性模型,441,656
contemporaneously exogenous	同期外生,318
continuous random variable	连续随机变量,667

control group	控制组,217
control variable	控制变量,23
controlling for too many factors in regression analysis	在回归分析中控制因素过多,196
corner solutions	角点解,529
corrected R-squared	调整 R-平方,193
correlation coefficient	相关系数,681,682
property CORR.1 and CORR.2	性质 CORR.1 和 CORR.2,682
correlation, no serial, Assumption TS.5′	假定 TS.5′,无序列相关,355
count variable	计数变量,546
county crime rates in North Carolina, Example 13.9	例 13.9,北卡罗莱纳州县一级的犯罪率,432
covariance	协方差,680
property COV.1, COV.2 and COV.3	性质 COV.1,COV.2 和 COV.3,681
covariance stationary	协方差平稳,348
process	～过程,348
covariate	协变量,23
crime, economic model of	犯罪的经济模型
Example 1.1	例 1.1,3
Example 5.3	例 5.3,173
Example 9.1	例 9.1,279
crime rate on clear-up rate, distributed lag of, Example 13.6	例13.6,犯罪率的滞后分布对清除率的影响,425
criteria, in-sample and out-of-sample	样本内与样本外准则,599
critical value	临界值,120,726
crop yield, effects of fertilizer on, Example 1.3	例 1.3,化肥对庄稼产量的影响 ,14
cross-sectional data	横截面数据 ,6
set on wages and other individual characteristics, Table 1.1	表 1.1,工资及其他个人特征的～集,7
cumulative distribution function(cdf)	累积分布函数(cdf) ,668

D

data	数据,630
censoring	截取～,551
collection	～搜集,619
cross-sectional	横截面～,6
entering and storing	进人和存储～,620
experimental	实验～,2
for Puerto Rico (minimum wage, unemployment, and related data), Table 1.3	表1.3,波多黎各(的最低工资、失业率和相关数据),9

frequency	频率,9
in WAGE1.RAW, partial listing, Table 7.1	表7.1,部分列出WAGE1.RAW中的～,212
inspecting, cleaning, and summarizing	～的检查、清理和总结,622
longitudinal	纵横～,408
mining	～开采,625
nonexperimental	非实验～,2
observational	观测～,2
on U.S. inflation and unemployment rates, partial listing of, Table 10.1	表10.1,部分列出美国通货膨胀和失业率～,312
panel	综列～,10
set on city crime statistics, two-year panel, Table 1.5	表1.5,城市犯罪统计的两年综列～集,12
set on economic growth rates and country characteristics, Table 1.2	表1.2,各国经济增长率和国家特征～集,8
structures, comment on	对～结构的评论,13
structure of economic	经济～的结构,5
time series	时间序列～,8
data bases, on-line	网上数据库,620
data scaling	数据测度
effects of, Table 6.1	表6.1,～的影响,179
on OLS statistics, effects of	～对OLS统计量的影响,178
data set, deciding on the appropriate	决定适当的数据集,619
Davidson-MacKinnon test	戴维森 麦金农检验,283
degrees of freedom (df)	自由度(df),57,99,693
demand for compact disks, Example A.2	例A.2,对密纹唱碟的需求,647
denominator degrees of freedom	分母自由度,143
dependent variable	因变量,23
inf, Table 12.2	表12.2,～inf,392
log($chnimp$), Table 12.1	表12.1,～log($chnimp$),391
log($wage$), Table 15.1	表15.1,～log($wage$),475
Table 9.2	表9.2,287
measurement error in	～中的测量误差,291
$narr86$, Table 9.1	表9.1,～$narr86$,280
predicting y when log(y) is the	当log(y)是～时预测y,202
$rprice$, Table 13.2	表13.2,～$rprice$,416
sav, Table 8.1	表8.1,～sav,264
Table 9.3	表9.3,290
dependent variable is log(y)	因变量是log(y)
interpreting coefficients on dummy explanatory variables when	当～时对虚拟解释变量系数的解释,218

predicting y when	当～时对 y 的预测,203
derivation of equation (3.22)	方程(3.22)的推导,109
derivation of the first order conditions, equations (3.13)	推导方程(3.13)的一阶条件,109
derivative	导数, 658
deriving the inconsistency in OLS	推导 OLS 的不一致性,165
descriptive statistic	描述静态,644
descriptive statistics	描述静态,643
determinants of college GPA	大学 GPA 的决定因素
Example 3.1	例 3.1,73
Example 3.4	例 3.4,80
Example 4.3	例 4.3,125
determinants of number of arrests for young men, Table 17.3	表 17.3,年轻人拘捕次数的决定因素,550
determinants of personal computer ownership, Example 8.9	例 8.9,个人拥有计算机的决定因素,273
determinants of women's fertility, Table 13.1	表 13.1,妇女生育能力的决定因素,411
detrending	除趋势,337
interpretation of regressions with a time trend	对含有时间趋势的回归的～解释,337
diagonal matrix	对角阵,745
Dickey-Fuller(DF) test	迪基-富勒(DF)检验,580
Dickey-Fuller distribution	迪基-富勒分布,580
difference-in-differences estimator	差异中的差分估计量,415
differences in slopes	斜率差异,226
differencing	差分
and serial correlation	～与序列相关,394
the interest rate equation, Example 12.6	例 12.6,利率方程～,394
with more than two time periods	多于两个时期的～,429
differential calculus	微积分,658
differentiation of linear and quadratic forms	线性和二次型的微分,751
diminishing marginal effect	边际效应递减,652
discrete random variable	离散随机变量,665
distributed lag models for housing investment, Table 18.1	表 18.1,住房投资的分布滞后模型,578
distributed lag of crime rate on clear-up rate, Example 13.6	例13.6,犯罪率的分布滞后对清除率的影响,425
distributions	分布
for the standardized estimators, Theorem 4.2	定理 4.2,标准化估计量的分布,117
normal and related	正态和相关分布,688
disturbance	扰动项,23,69

downward bias	向下的偏误,91
drunk driving laws on traffic fatalities, effect of, Example 13.7	例13.7,酒后驾驶法律对交通事故致死人数的影响,428
dummy explanatory variables when the dependent variable is log(y), interpreting coefficients on	当因变量是 log(y)时对虚拟解释变量系数的解释,218
dummy variable regression	虚拟变量回归,445
dummy variable trap	虚拟变量陷阱,213
dummy variables	虚拟变量,211,324
for multiple categories	多个虚拟变量,220
incorporating ordinal information by using	通过使用～而包含序数信息 ,221
interactions involving	涉及～的交互作用 ,224
duration analysis	久期分析, 553
duration of recidivism, Example 17.4	例 17.4,累犯的相隔期间, 553
Durbin-Watson (DW) statistic	德宾-沃森统计量,382
Durbin-Watson test under classical assumptions	古典假定下的德宾-沃森检验, 382
dynamically complete model	动态完备模型,368
and the absence of serial correlation	动态完备模型和无序列相关, 366
E	
econometric analysis	计量经济分析,623
causality and the notion of ceteris paribus in	～中的因果性和其他条件不变的概念,13
econometric model	计量经济模型,4
and estimation methods	～与估计方法,627
econometrics	计量经济学,1
economic analysis, steps in empirical	实证经济分析的步骤,2
economic data, structure of	经济数据的结构, 5
economic growth rates and country characteristics data set on, Table 1.2	表 1.2,经济增长率和国家特征数据集,8
economic model	经济模型,2
economic model of crime	犯罪的经济模型
Example 1.1	例 1.1,3
Example 5.3	例 5.3,173
Example 9.1	例 9.1,279
economic performance, election outcomes and, Example 10.6	例 10.6,选举结果和经济表现,330
economic significance	经济显著性,131
education on fertility, effect of, Example 15.9	例 15.9,教育对生育力的影响,489
education	教育
estimated relationship between the probability of being in the labor force and years of, Fig-	图7.3,估计参与劳动市场的概率与受～年数之间的关系,235

ure 7.3	
for married women, estimating the return to, Example 15.1	例 15.1,估计已婚妇女受～的回报,467
for men, estimating the return to, Example 15.2	例 15.2,估计男性受～的回报,468
wage and, Example 2.4	例 2.4,工资与～,34
Example 2.7	例 2.7,38
efficiency	有效性,707
and inference	～与推断,377
of OLS: the Gauss-Markov theorem	OLS 的～:高斯-马尔科夫定理,101
efficient markets hypothesis	有效市场假说
Example 11.4	例 11.4,356
heteroskedasticity and, Example 12.8	例 12.8,异方差性与～,400
elasticity	弹性,46,655
election outcomes and economic performance, Example 10.6	例 10.6,选举结果与经济表现,330
empirical analysis	经验分析,2
empirical paper	经验论文
conceptual (or theoretical) framework	概念(或理论)框架,627
conclusion	结论,632
data	数据,630
econometric models and estimation methods	计量模型和估计方法,627
introduction	引言,626
results	结论,631
style hints	风格提示,632
writing	写作,626
employment and the minimum wage, Puerto Rican, Example 10.3	例 10.3,波多黎各的就业和最低工资,324
employment, Puerto Rican, Example 10.9	例 10.9,波多黎各的就业,338
endogeneity	内生性
and testing overidentifying restrictions, testing for	～检验和对过度识别约束的检验,482
of a single explanatory variable	单一个解释变量的～, 84,278,461
endogenous sample selection	内生样本选择,299
endogenous variables	内生变量,503
Engle-Granger two-step procedure	恩格尔-葛兰杰两阶段程序,593
enrollment, campus crime and, Example 4.4	例 4.4,校园犯罪与注册人数,126
enterprise zones on business investments, effect of	企业特区对商业投资的影响,727
enterprise zones on unemployment claims, effect of, Example 13.8	例 13.8,企业区对失业公开数据的影响,431

equations, systems with more than two	两个以上方程的方程组,514
error correction model	误差纠正模型,586,591,592
for holding yields, Example 18.7	例 18.7,持有收益的~,592
error term	误差项,23,69
error variance	误差方差,53,94
adding regressors to reduce	增加回归元来减少~,197
estimating	估计~,56
errors-in-variables	变量中误差,461
problems, IV solution to	~问题的 IV 解法,480
estimate	估计值,701
estimation	估计,515
by 2SLS	2SLS~,512
estimators	估计量,701
asymptotic or large sample properties of	~的渐近或大样本性质,708
bias of	~的偏误,703
consistent	~的一致性,709
finite sample properties of	~的有限样本性质,700
least squares	最小二乘~,715
maximum likelihood	最大似然~,714
minimum variance unbiased	最小方差无偏~,715
sampling distribution of two unbiased, Figure C.2	图 C.2,两个无偏~的抽样分布,705
sampling variance of	~的抽样方差,704
unbiased, Figure C.1	图 C.1,无偏~,703
event study	事件研究,327
Cochrane-Orcutt estimation in, Example 12.4	例 12.4,~中的科克伦-奥克特估计,390
excluding a relevant variable	排除一个有关变量,87
exclusion restrictions	排除性约束,140,476, 510
testing	检验~,139
exogenous explanatory variables	外生解释变量 ,84,474
exogenous instrumental variables, Assumption 2SLS.3	假定 2SLS.3,外生工具变量,498
exogenous sample selection	外生样本选择,299,599
exogenous variables	外生变量,472.503
expectations augmented Philips curve, Example 11.5	例 11.5,预期增广菲利普斯曲线,357
expectations hypothesis, Example 1.7	例 1.7,期望假设,18
expected revenue, finding, Example B.5	例 B.5,求预期收益,675
expected value	期望值,672,527
computing, Example B.3	例 B.3,计算~,673

Figure B.5	图 B.5,685
of χ^2, Example B.4	图 B.4,χ^2 的～,674
properties of	～的性质,674,752
property E.1 and E.2	性质 E.1 和 E.2,674
property E.3	性质 E.3,675
experiment	实验,664
experimental data	实验数据,2
experimental group	实验组,217
explained sum of squares (SSE)	解释平方和(SSE),38,78
explained variable	解释变量,23
in equation (10.3), example of X for, Table 10.2	表 10.2,方程(10.3)～X 的例子,317
measurement error in	～中的测量误差,294
using proxy variables for unobserved	使用观测不到的～的代理变量,284
exponential function	指数函数,657
exponential smoothing	指数平滑,594
exponential trend	指数趋势,333
F	
F and *t* statistics, relationship between	*F* 和 *t* 统计量之间的关系,145
F distribution	*F* 分布,694,753
Figure B.11	图 B.11,696
property of	～的性质,753
F random variable	*F* 随机变量,753
F statistic	*F* 统计量,142
for overall significance of a regression	作为一个回归总体显著性的～,148
F tests, computing *p*-values for	计算 *F* 检验的 *p* 值,147
testing multiple linear restrictions	检验多个线性约束的～,139
family saving equation, Example 8.6	例 8.6,家庭储蓄方程,264
feasible GLS (FGLS)	可行 GLS(FGLS),389
estimation of the AR(1) model	对 AR(1)模型的～估计,389
estimation with AR(1) errors	含有 AR(1)误差的～估计,389
estimator (FGLS)	～估计量,(FGLS),267
heteroskedasticity function must be estimated	必须估计的异方差性函数,266
procedure to correct for heteroskedasticity	修正异方差性的程序,268
with heteroskedasticity and AR(1) serial correlation	含有异方差性和 AR(1)序列相关的～,403
fertility and personal exemption, cointegration between, Example 18.5	例18.5,生育率和个人税收豁免之间的协积
Example 10.8	例 10.8,336

Example 11.6	例 11.6,364
Example 11.8	例 11.8,368
fertilizer on crop yield, effects of, Example 1.3	例 1.3,肥料对作物产量的影响,14
fertilizer, soybean yield and, Example 2.1	例 2.1,大豆产量与肥料,24
finite distributed lag (FDL) model	有限分布滞后(FDL)模型,313
Example 11.2	例 11.2,353
finite sample properties	有限样本性质
of estimators	估计量的～,700
of OLS	OLS 的～,758
under classical assumptions	经典假定下～,316
firm performance, CEO compensation and, Example 6.4	例 6.4,CEO 薪酬与企业业绩,195
firm sales, CEO salary and, Example 2.11	例 2.11,CEO 薪水与企业销售量,44
firm size, R&D intensity and, Example 9.8	例 9.8,R&D 强度与企业规模,301
first difference	一阶差分,363
first differencing, fixed effects or	固定效应或一阶差分, 447
first order conditions	一阶条件,31,72,756
first-differenced equation	一阶差分方程,422
fitted value	拟合值,30
fitted values and residuals	拟合值和残差, 36
Figure 2.4	图 2.4,31
Table 2.2	表 2.2,36－37
5% critical value and rejection region, Figure 4.7	图 4.7,5%的临界值和拒绝域,144
5% rejection rule	5%的拒绝法则
Figure 4.2	图 4.2,119
Figure 4.3	图 4.3,122
Figure 4.4	图 4.4,124
fixed effects	固定效应,420
estimator	估计量
model	～模型,420
or first differencing	～或一阶差分,447
random effects or	随机效应或～,452
with unbalanced panels	具有不平衡综列数据的～,448
fixed effects estimation	固定效应估计,441
of the scrap rate equation, Table 14.1	表 14.1,对废弃率方程的～,443
forecast	预测
comparing one step ahead	与提前一步～相比,599
conditional and unconditional	条件和无条件～,595
error～	误差
for the unemployment rate, two-year-ahead,	例 18.10,提前两年对失业率的～,603

Example 18.10	
interval	—区间,596
multiple step	多步～,601
out-of-sample comparisons of unemployment, Example 18.9	例 18.9,样本(范围)外失业比较,600
point	督促～,596
forecasting	预测,593
one-step-ahead	提前一步～,596
the U.S. unemployment rate, Example 18.8	例 18.8,美国失业率～,596
trending, seasonal, and integrated processes	趋势、季节性和单整过程～, 603
types of regression models used for	几种用于～的回归模型, 595
401 (k) pension plans, participation in, Example 3.3	例 3.3,401 k养老金计划的参与,78
401 (k) plans, participation rates in, Example 4.6	例 4.6,401 k养老金计划的参与率,131
free throw shooting	罚球投篮
Example B.1	例 B.1,670
Example B.2	例 B.2,672
function	函数
population regression (PRF)	总体回归(PRF)～ , 26
sample regression (SRF)	样本回归(SRF) ～, 72
functional form	函数形式,183, 324
involving logarithms, summary of, Table 2.3	表 2.3,对涉及对数的～的总结, 45
units of measurement and	测量单位与～, 41
functional form misspecification	函数形式误设,278
RESET as a general test for	RESET 作为对～的一般检验, 281

G

Gauss-Markov assumptions	高斯-马尔科夫假定,93
Gauss-Markov theorem	高斯-马尔科夫定理,101
Theorem 10.4	定理 10.4,322
Theorem 3.4	定理 3.4,101
Theorem E.3	定理 E.3,760
variances of the OLS estimators and	OLS 估计量和～的方差, 320
generalized least squares (GLS) estimators	广义最小二乘(GLS)估计量,262
geometric (or Koyck) distributed lag	几何(或考依克)分布滞后, 574
goodness-of-fit	拟合优度,39,78
and selection of regressors	～与回归元的选择,192
GPA equation with measurement error, Example 9.7	例 9.7,含有测量误差的 GPA 方程,296

Granger causality	葛兰杰因果关系,598
graphs	图
crime, Figure 4.5	图 4.5 的～*crime*,127
equation (7.16),Figure 7.2	方程(7.16)的图 7.2,226
$E(Y\|x)$,Figure B.6	$E(Y\|x)$的图 B.6,686
logistic function, Figure 17.1	逻辑斯蒂函数图 17.1, 531
quantity, Figure A.2	数量图 A.2, 648
wage, Figure 7,1	工资图 7,1,214
y	y 图
Figure A.3	图 A.3,651
Figure A.4	图 A.4,654
Figure A.5	图 A.5,657
y_t, Figure 2.7	y_t 图 2.7,48
growth rate	增长率,334

H

Heckit method	赫克曼方法,561
hedonic price model for houses, Example 4.8	例 4.8,住房的按质论价模型,135
heterogeneity bias	异质性偏误,421
heteroskedasticity	异方差,54
and serial correlation in regression models	回归模型中的～和序列相关,402
and the efficient markets hypothesis, Example 12.8	例 12.8,～与有效市场假说,400
Breusch-Pagan test for	对～的布罗施-培甘检验, 257
feasible GLS procedure to correct for	校正～的可行的 GLS 程序,268
for OLS, consequences of	～对 OLS 造成的后果,248
function must be estimated: feasible GLS	必须估计的～函数:可行的 GLS,266
in a wage equation, Example 2.13	例 2.13,工资方程中的～,54
in housing price equations, Example 8.4	例 8.4,住房价格方程中的～,258
in time series regressions	时间序列回归中的～,398
is known up to a multiplicative constant	除一个常数倍数外都是已知的～,261
of unknown form	未知形式的～,249
special case of the White test for	～怀特检验的特殊情形, 260
testing for	对～的检验,255,399
2SLS with	有～的 2SLS,486
White test for	对～的怀特检验, 259
heteroskedasticity-robust F statistic	异方差 稳健 F 统计量, 252
Example 8.2	例 8.2,253
heteroskedasticity-robust inference after OLS estimation	OLS 估计后的异方差-稳健推断,249

heteroskedasticity-robust *LM* statistic	异方差-稳健 *LM* 统计量,253,254
Example 8.3	例 8.3,254
heteroskedasticity-robust *LM* tests	异方差-稳健 *LM* 检验, 253
heteroskedasticity-robust standard error	异方差-稳健标准误, 251
log wage equation with, Example 8.1	例 8.1,具有~的对数工资方程,251
heteroskedasticity-robust statistics	异方差-稳健统计量, 399
heteroskedasticity-robust *t* statistic	异方差-稳健 *t* 统计量,251
highly persistent	高度持续性,358－359
highly persistent time series	高度持续性时间序列
in regression analysis	回归分析中的~,358
transformations on	~的变换, 362
hiring, race discrimination in	雇用中的种族歧视
Example C.3	例 C.3,723
Example C.5	例 C.5,730
Example C.7	例 C.7,734
histogram of *prate*, Figure 5.2	图 5.2,*prate* 的直方图,168
holding yields, error correction model for, Example 18.7	例18.7,持有收益率的误差纠正模型,592
homoskedastic normal distribution with a single explanatory variable, Figure 4.1	图4.1,具有单一个解释变量的同方差正态分布,115
homoskedasticity	同方差性, 52
and no serial correlation, Assumption E.4	假定 E.4,~与无序列相关,759
assumption for time series models	时间序列模型的~假定, 369
Assumption MLR.5	假定 MLR.5, 93
Assumption SLR.5	假定 SLR.5, 52
Assumption TS.4	假定 TS.4, 320
Assumption TS.4′	假定 TS.4′, 355
Assumption 2SLS.5	假定 2SLS.5, 490
hourly wage equation	小时工资方程
Example 3.2	例 3.2, 74
Example 3.6	例 3.6, 90
Example 4.1	例 4.1, 120
Example 7.1	例 7.1, 215
housing expenditures and saving, Example 16.2	例 16.2,住房支出与储蓄,505
housing, graph of, Figure A.1	图 A.1,*housing* 图, 647
housing investment	住房投资
and prices, Example 10.7	例 10.7,住房投资与价格, 335
and residential price inflation, Example 18.1	例18.1,住房投资与住房价格的通货膨胀,577
distributed lag models for, Table 18.1	表 18.1,住房投资的分布滞后模型, 578

Example 10.10	例 10.10,339
housing price equation	住房价格方程
Example 9.2	例 9.2,282
heteroskedasticity in, Example 8.4	例 8.4,～中的异方差性, 258
housing price regression, Example 7.4	例 7.4,住房价格回归,218
housing prices	住房价格
and air pollution, Example 4.5	例 4.5,～与空气污染, 128
and distance from an incinerator, Example 5.1	例 5.1,～与距垃圾焚化炉的距离,166
effect of a garbage incinerator's location on, Example 13.3	例13.3,垃圾焚化炉的位置对～的影响,414
effects of pollution on, Example 6.1	例 6.1,污染对～的影响, 182
Example 6.2	例 6.2,188
hypothesis	假设
about β_j, testing other	检验对 β_j 的其他～, 125
about the mean in a normal population, testing	检验对一个正态总体均值的假设,725
efficient markets, Example 11.4	例 11.4,市场有效性～, 356
expectations, Example 1.7	例 1.7,预期～,18
test	～检验, 724
testing multiple	检验多重～, 534
hypothesis testing	假设检验, 724
fundamentals of	～基础知识, 724
relationship between confidence intervals and	置信区间与～之间的关系,734

I

idempotent matrix	幂等矩阵, 750
properties of	～的性质, 751
identification	识别, 465
in a two-equation system	两方程联立方程组中的～, 508
in systems with three or more equations	两个以上方程的方程组中的～,514
of a structural equation, rank condition for	结构方程～的秩条件, 510
of an equation, order condition for	方程～的阶条件,480
order condition for	～的阶条件, 515
identified equation	可识别方程, 508
identity matrix	单位矩阵, 745
idiosyncratic error	特异误差,420
impact multiplier	冲击(即期)乘数,314
impact propensity	即期倾向,314
in-sample criteria	样本内准则, 599
incidental truncation	偶然断尾, 577,560
inclusion of an irrelevant variable	包含一个无关变量, 87

inconsistency	不一致性，165
inconsistent	不一致的，709
independence	独立，669
independent random variables	独立随机变量，670
independent variable	自变量，23
changing more than one simultaneously	同时改变不止一个～，75
linear relationships among	～之间的线性关系，95
independently pooled cross section	独立混合横截面，408
index number	指数，324，327
inference	推断
efficiency and	有效性与～，377
under the classical linear model assumption	在经典线性模型假定下的～，322
infinite distributed lag (IDL) model	无限分布滞后(IDL)模型，572
inflation and deficits on interest rates, effects of, Example 10.2	例10.2,通货膨胀和赤字对利率的影响,323
inflation and openness	通货膨胀与开放
Example 16.4	例 16.4，512
Example 16.6	例 16.6，514
inflation and unemployment rates, partial listing of data on U.S., Table 10.1	表10.1,部分列举美国通货膨胀和失业率的数据,312
influential observations	有重大影响的观测，300
information set	信息集，593
instrumental variables (IV)	工具变量(IV),461
estimator	～估计量,465
estimation, computing R-squared after	在～估计后计算 R-平方，471
estimation of the multiple regression model	多元回归模型的～估计，472
for education, using college proximity as, Example 15.4	例 15.4,以大学毗邻作为教育的～，474
solutions to errors-in-variables problems	变量中误差问题的～解法，480
interaction effect	交互效应，190
interaction term	交互作用项,224
intercept	截距，69，646
parameter	～参数，23
shift	～漂移，213
interest rates	利率
cointegrating parameter for	～的协积参数,591
effects of inflation and deficits on, Example 10.2	例 10.2,通货膨胀和赤字对利率的影响,323
internet	互联网，617
interval estimation	区间估计，715

interval estimator	区间估计量，716
inverse	逆，749
Mills ratio	～米尔斯比率，542
properties of	～的性质，749
IQ as a proxy for ability, Example 9.3	例 9.3，*IQ* 作为能力的一个代理变量，286
irrelevant variable, inclusion of	包含无关变量，87
iterated expectations, law of	重叠期望法则，687

J

job training and worker productivity	在职(工作)培训和工人生产率
Example 1.2	例 1.2，4
Example 15.10	例 15.10，489
job training grants on firm scrap rates, effect of, Example 4.7	例4.7，在职(工作)培训津贴对企业废弃率的影响，132
job training grants on worker productivity, effect of	在职(工作)培训津贴对工人生产率的影响
Example C.2	例 C.2，720
Example C.6	例 C.6，732
job training on firm scrap rates, effect of	在职(工作)培训对企业废弃率的影响
Example 14.1	例 14.1，443
Example 14.3	例 14.3，448
joint and conditional distributions, features of	联合和条件分布的性质，680
joint distribution	联合分布，669
joint hypotheses test	联合假设检验，140
jointly insignificant	联合不显著，145
jointly statistically significant	联合统计显著，144
just identified equation	恰好可识别方程，515

L

labor force participation	劳动市场参与
LPM, logit, and probit estimates of, Table 17.1	表 17.1，～的 LPM，logit 和 probit 估计值，538
married women's, Example 17.1	例 17.1，已婚妇女的～，537
of married women, Example 8.8	例 8.8，已婚妇女的～，271
labor supply	劳动供给
function, Example A.7	例 A.7，～函数，657
married women's annual, Example 17.2	例 17.2，已婚妇女的年～，543
labor supply of married, working women	已婚工作妇女的劳动供给
Example 16.3	例 16.3，510
Example 16.5	例 16.5，512

lag distribution	分布滞后, 314
for the rational distributed lag (18.16), Figure 18.1	图 18.1,理性滞后分布(18.16)的~, 577
with two nonzero lags, Figure 10.1	图 10.1,具有两个非零滞后的~, 315
lagged dependent variable	滞后因变量, 289
serial correlation in the presence of	出现~时的序列相关, 378
lagged endogenous variable	滞后内生变量, 517
Lagrange multiplier *LM* statistic	拉格朗日乘数 *LM* 统计量, 171
for q exclusion restrictions	q 个排除性约束的~, 172
other large sample tests	其他大样本检验,171
large sample properties	大样本性质, 162
latent variable model	潜变量模型, 532
law enforcement on city crime levels, effect of, Example 1.5	例1.5,执法力度对城市犯罪水平的影响, 16
law of iterated expectations	重期望法则,687
law of large numbers (LLN)	大数定律, 710
law school rankings on starting salaries, effects of, Example 7.8	例 7.8,法学院排名对起薪的影响,223
leads and lags estimator	前导和滞后估计量, 591
least absolute deviations (LAD)	最小绝对离差法, 304
least squares	最小二乘, 715
estimator	~估计量,715
likelihood ratio statistic	似然比统计量,535
limited dependent variable (LDV)	限值因变量,529
models, asymptotic standard errors in	~模型中的渐近标准误,569
linear and quadratic forms, differentiation of	线性和二次型的微分, 751
linear function	线性函数, 646
Figure 2.1	表 2.1,26
properties of	~的性质, 645
linear housing expenditure function, Example A.1	例 A.1,线性住房支出函数, 646
linear in parameters	对参数是线性的
Assumption E.1	假定 E.1, 758
Assumption MLR.1	假定 MLR.1,82
Assumption SLR.1	假定 SLR.1,47
Assumption TS.1	假定 TS.1,47
Assumption 2SLS.1	假定 2SLS.1,498
linear independence	线性独立, 749
linear probability model (LPM)	线性概率模型,233
a binary dependent variable	二值因变量的~, 232
by weighted-least squares, estimating	用加权最小二乘法估计~, 272

of arrests, Example 7.12	例 7.12,拘捕次数的～,236
revisited	重访的～,271
linear regression	线性回归, 46
linear relationships among the independent variables	自变量之间的线性关系,95
linear restrictions, testing general	检验一般的线性约束, 149
linear time trend	线性时间趋势, 332
linear trend line, Figure 18.2	图 18.2,线性趋势线, 604
linear unbiased estimator in the AR(1) model, obtaining the best	在 AR(1)模型中得到最优线性无偏估计量,387
linearity and weak dependence, Assumption TS, 1′	假定 TS,1′,线性和弱依赖,352
linearly independent vectors	线性独立向量, 749
literature review	文献回顾, 618
log function	对数函数,653
log hourly wage equation	对数小时工资方程
Example 7.10	例 7.10,227
Example 7.5	例 7.5,219
Example 7.6	例 7.6,220
log of U.S. real GDP, unit root in, Example 18.4	例18.4,美国真实 GDP 的对数中的单位根,583
log wage equation	对数工资方程
Example 2.10	例 2.10,44
with heteroskedasticity-robust standard errors, Example 8.1	例 8.1,具有异方差稳健标准误的～,251
log(*price*) as a quadratic function, Figure 6.2	图 6.2,log(*price*)作为一个二次函数,189
log-likelihood	对数似然, 534
logarithmic functional forms	对数函数形式, 183
logarithmic wage equation, Example A.6	例 A.6,对数工资方程,656
logit and probit estimates, interpreting	解释 logit 和 probit 的估计值, 535
logit and probit models	logit 和 probit 模型
for binary response	二值响应的～,530
maximum likelihood estimation of	～的最大似然估计, 533
specifying	设定～, 530
logit model	logit 模型,530
long-run elasticity	长期弹性,325
long-run multiplier	长期乘数,314
log-run propensity (LRP)	长期倾向,314
longitudinal data	纵横数据,408
loss function	损失函数,593

LPM, logit, and probit estimates of labor force participation, Table 17.1	表 17.1,劳动市场参与的 LPM,logit 和 probit 估计值,538
M	
MA(1)	一阶移动平均 MA(1) ,350
manufacturing firms, scrap rates for, Table C.3	表 C.3,制造业的废弃率,720－721
marginal effect	边际效应,646
married women's annual labor supply, Example 17.2	例 17.2,已婚妇女的年度劳动供给,543
married women's labor force participation, Example 17.1	例 17.1,已婚妇女的劳动市场参与,537
martingale	鞅,594
difference sequence	～差分序列, 579
matched pair samples	配对样本,454
matrix	矩阵,744
addition	～加法, 745
diagonal	对角阵, 745
idempotent	幂等矩阵,750,751
identify	单位矩阵,745
multiplication	～乘法, 746,747
notation	～符号,756
operations	～运算,745
properties of positive definite and positive semidefinite	正定和半正定～的性质,750
rank of	～的秩, 749
square	方阵, 744
zero	零～,745
maximum likelihood	最大似然,714
estimation of logit and probit models	logit 和 probit 模型的～估计,533
estimator	～的估计量,714
mean absolute error	平均绝对误差,600
mean squared error(MSE)	均方误(MSE),708
measure of central tendency	中心趋势的度量
expected value	～的期望值, 672
the median	～的中位数,676
measurement	测量
and functional form, units of	～单位和函数形式,41
on OLS statistics, effects of changing units of	改变～单位对 OLS 统计量的影响, 41
measurement error	测量误差, 278
GPA equation with, Example 9.7	例 9.7,含～的 GPA 方程,296

in an explanatory variable	一个解释变量中的～,294
in scrap rates, Example 9.6	例 9.6,废弃率中的～,293
in the dependent variable	因变量中的～,291
properties of OLS under	有～时 OLS 的性质,291
saving function with	含～的储蓄函数, 292
measures of association: covariance and correlation	关联的度量:协方差和相关, 680
measures of variability: variance and standard deviation	变异的度量:方差和标准差,676
measuring the return to education, Example 1.4	例 1.4,对教育回报的度量,15
mechanics and interpretation of ordinary least squares	普通最小二乘法的机理和解释,71
median	中位数, 645,676
method of moments	矩法,713
micronumerosity	微数缺测性,96
minimum variance unbiased	最小方差无偏,763
estimator	～的估计量 114,715
minimum wage	最低工资
on unemployment, effect of, Example 1.6	例 1.6,最低工资对失业的影响,17
Puerto Rican, Example 12.7	例 12.7,波多黎各,398
Puerto Rican employment and, Example 10.3	例 10.3,波多黎各的就业与～,324
unemployment, and related data for Puerto Rico, Table 1.3	表 1.3,波多黎各的～、失业和有关数据, 9
missing data	缺失数据, 298
misspecification analysis	误设分析, 88
misspecified models, variances in	误设模型中的方差, 97
model	模型
and ordinary least squares estimation	～和普通最小二乘估计, 755
AR(1), Example 11.3	例 11.3,AR(1)～, 354
binary response	二值响应～,233
bivariate linear regression	双变线性回归～,22
censored regression	截取回归～, 551
classical linear	经典线性～,114
constant elasticity	常弹性～, 44, 656
dynamically complete	动态完全～, 368
econometric	计量经济～,4
economic	经济～, 2
error correction	误差纠正～,591
finite distributed lag (FDL)	有限分布滞后～, 313
Example 11.2	例 11.2, 353
fixed effects	固定效应～,420

infinite distributed lag (IDL)	无限分布滞后～, 572
latent variable	潜变量～, 532
linear probability (LPM)	线性概率～, 233
logit	logit～, 530
multiple linear regression	多元线性回归～, 69
nonnested	非嵌套～, 194
Poisson regression	泊松回归～, 546
population	总体～, 82
probit	probit～, 531
random effects	随机效应～, 449
rational distributed lag (RDL)	理性分布滞后(RDL)～, 576
restricted	受约束～, 141
simple linear regression	简单线性回归～, 22
simple regression	简单回归～, 22
static	静态～, 313
Example 11.1	例 11.1, 353
Tobit	Tobit～, 540
true	真实～, 82
truncated regression	断尾回归～, 551
two-variable linear regression	两变量线性回归～, 22
underspecifying	设定不足～, 87
unobserved effects	不可观测效应～, 420
unrestricted	无约束～, 141
vector autoregressive (VAR)	向量自回归(VAR)～, 598
with interaction terms	含交互项的～, 190
with k independent variables	含有 k 个自变量的～, 69
with quadratics	含二次项的～, 185
with two independent variables	含两个自变量的～, 66
moments and distributions of random vectors	随机向量的矩和分布, 751
moving average process of order one [MA(1)]	一阶移动平均[MA(1)], 350
multicollinearity	多重共线性, 95
and 2SLS	～和 2SLS, 479
components of the OLS variances	OLS 方差中～的成分, 94
multiple regression	多元回归
a "partialling out" interpretation of	～"排除其他因素影响"的解释, 76
analysis	～分析, 66
estimates, comparison of simple and	简单和～估计值的比较, 77
motivation for	～的动因, 66
on the meaning of "holding other factors fixed" in	～中"保持其他因素不变"的意义, 75

terminology for, Table 3.1	表 3.1, ～术语表, 70
multiple restrictions	多重约束, 140
multiple step forecasts	多步预测, 594, 601
multiplicative measurement error	乘数测量误差, 293
multivariate normal distribution	多变量正态分布, 752
properties of	～的性质, 753
murder rates and size of the police force, Example 16.1	例 16.1, 谋杀率和警力规模, 504

N

n-*R* squared statistic	*n*-*R*-平方统计量, 172
natural experiment	自然实验, 417, 469
natural logarithm	自然对数, 653
no perfect collinearity	非完全共线性
Assumption E.3	假定 E.3, 758
Assumption MLR.4	假定 MLR.4, 84
Assumption TS.3	假定 TS.3, 319
Assumption TS.3'	假定 TS.3', 352
no serial correlation	无序列相关
Assumption TS.5	假定 TS.5, 320
Assumption TS.5'	假定 TS.5', 355
Assumption 2SLS.6	假定 2SLS.6, 500
nominal U.S. imports, Figure 10.3	图 10.3, 美国名义进口量, 333
nonexperimental data	非实验数据, 2
nonlinear function	非线性函数, 650
nonlinearities in simple regression, incorporating	简单回归中引入非线性, 42
nonnested alternatives, tests against	对非嵌套对立假设的检验, 283
nonnested models	非嵌套模型, 194, 283
nonnormal populations	非正态总体
asymptotic confidence intervals for	～的渐近置信区间, 722
asymptotic tests for	～的渐近检验, 729
nonrandom sample	非随机样本, 298, 299
selection	～选择, 557
nonstationary process	非平稳过程, 348
normal and related distributions	正态和相关分布, 688
normal distribution	正态分布, 688, 689
additional properties of	～的额外性质, 692
simulated confidence intervals from, Table C.2	表 C.2, 模拟～的置信区间, 717-718
simulation of estimators for, Table C.1	表 C.1, 模拟～的估计量, 706-707
normal probability density function, Figure B.7	图 B.7, 正态概率密度函数, 689

normal random variable, probabilities for, Example B.6	例 B.6,正态随机变量的密度, 692
normal sampling distributions	正态抽样分布
Theorem 10.5	定理 10.5, 322
Theorem 4.1	定理 4.1, 115
normality	正态性
assumption	～假定,113
Assumption MLR.6	假定 MLR.6, 113
Assumption TS.6	假定 TS.6,322
of $\hat{\beta}$,Theorem E.5	定理 E.5,$\hat{\beta}$ 的～,762
of errors, Assumption E.5	假定 E.5,误差的正态性,761
notation, remarks on	对记号的评论, 736
null hypothesis	虚拟假设,117,724
number of arrests	拘捕次数
for young men, determinants of, Table 17.3	表 17.3,年轻人拘捕次数的决定因素,550
Poisson regression for, Example 17.3	例 17.3,拘捕次数的泊松回归,549
numerator degrees of freedom	分子自由度, 143

O

observational data	观测数据, 2
OLS (ordinary least squares)	普通最小二乘(法), 30,71
and FGLS, comparing	～和 FGLS 的比较, 390
and Tobit estimation of annual hours worked, Table 17.2	表 17.2,对年工作小时数的～和 Tobit 估计, 544
asymptotic properties of	～的渐近性质, 352
consequences of heteroskedasticity	异方差性对～造成的后果, 248
deriving the inconsistency in	推导～中的不一致性, 165
estimates	～估计值,27,71
fitted values and residuals	拟合值与残差, 75
intercept estimate	截距估计值, 72
mechanics and interpretation of	～的操作与解释, 35,71
on the selected sample consistent	对选择样本一致的～, 558
regression equation, interpreting	对～回归方程的解释, 72
results, Table 3	表 3,～的结果,634
sampling variances, Theorem 10.2	定理 10.2,～抽样方差,321
serial correlation-robust inference after	～后序列相关-稳健的推断, 395
simultaneity bias in	～中的联立偏误, 506
slope estimates	～斜率估计值,72
slope estimators, sampling variances of, Theorem 3.2	定理 3.2,～斜率估计量的抽样方差,93

the Gauss-Markov theorem, efficiency of	～的有效性:高斯-马尔科夫定理,101
under classical assumptions, finite sample properties of	～在经典假定下的有限样本性质，316
under measurement error, properties of	～在有测量误差时的性质，291
variances: multicollinearity, components of	～方差中多重共线性的成分，94
with serially correlated errors, properties of	有序列相关误差的～的性质，376
OLS, asymptotic efficiency of	OLS 的渐近有效性，173
Theorem 5.3	定理 5.3，175
OLS, asymptotic normality of	OLS 的渐近正态性
Theorem 11.2	定理 11.2,355
Theorem 5.2	定理 5.2,168
OLS, consistency of	OLS 的一致性
Theorem 11.1	定理 11.1,353
Theorem 5.1	定理 5.1，164
OLS, estimation	OLS 估计
heteroskedasticity-robust inference after	～后异方差-稳健推断，249
model and	模型和～,755
OLS estimators	OLS 估计量
and the Gauss-Markov theorem, variances of	～的方差和高斯-马尔科夫定理，320
estimating σ^2; standard errors on	估计 σ^2:～的标准误,99
expected value of	～的期望值，82
and variances of	～的方差,46
sampling distributions of	～的抽样分布，113
sampling variances of, Theorem 2.2	定理 2.2,～的抽样方差,55
variance of	～的方差,52,92
OLS regression line	OLS 回归线，31，72
and the population regression function, Figure 2.5	图 2.5,～和总体回归函数,33
OLS statistics	OLS 统计量
algebraic properties of	～的代数性质,37
effects of changing units of measurement on	改变度量单位对～的影响,41
effects of data scaling on	数据尺度对～的影响，178
OLS, unbiasedness of	OLS 的无偏性，46,316
Theorem 10.1	定理 10.1,319
Theorem 2.1	定理 2.1,50
Theorem 3.1	定理 3.1，86
omitted variable bias	遗漏变量偏误，89
more general cases	更一般的情形，91
the simple case	简单情形,87
omitted variables	遗漏变量，461

in a simple regression model	简单回归模型中的～, 462
on-line data bases	在线数据库, 620
on-line search services	在线搜索服务, 619
one-sided alternative	单侧对立假设, 118, 726
p-value, Figure C.7	图 C.7, p 值, 731
Figure C.8	图 C.8, 733
rejection region against, Figure C.5	图 C.5、～的拒绝域, 727
testing against	相对～的检验, 118
one-tailed test	单尾检验, 120, 727
order condition	阶条件, 480, 510
for identification	识别的～, 515
of an equation	识别一个方程的～, 480
ordinal information by using dummy variables, incorporating	通过虚拟变量包含序数信息, 221
ordinal variable	序数变量, 222
ordinary least squares, see OLS	普通最小二乘法, 见 OLS
organizing panel data	编排综列数据, 425
out-of-sample comparisons of unemployment forecasts, Example 18.9	例 18.9, 失业率预测的样本外比较, 600
out-of-sample criteria	样本外准则, 599
outliers	异常数据, 300
outlying observations	异常观测, 298, 300
output per labor hour in the United States, Figure 10.2	表 10.2, 美国每小时劳动的产出, 322
overall significance of the regression	回归的整体显著性, 149
overdispersion	过度分散, 548
overidentification restrictions	过度识别约束, 484
testing	检验～, 484, 485
overidentified equation	过度识别的方程, 515
overidentifying restrictions, testing for endogeneity and testing	检验内生性和检验过度识别约束, 482
overspecifying the model	过度设定模型, 87

P

p-value	p 值, 129, 730
computing and using	计算和使用～, 730
Figure 4.6	图 4.6, 130
for t tests, computing	用于 t 检验计算～, 128
one-sided alternative, Figure C.7	图 C.7, 单侧对立假设的～, 731
Figure C.8	图 C.8, 733

summary of how to use	如何使用～概要,734
pairwise uncorrelated random variables	两两不相关随机变量,683
panel(longitudinal) data	综列(纵横)数据,10
applying 2SLS to pooled cross sections and	利用 2SLS 混合横截面数据和～,488
methods to other data structures applying	适用～的其他数据结构方法,453
simultaneous equation models with	利用～的联立方程模型,520
wage equation using, Example 14.4	例 14.4,利用～的工资方程,451
panel data set	综列数据集,408
parameter	参数,699
estimation, general approaches to	～估计的一般方法,713
intercept	截距～,23
slope	斜率～,23
parents' education in a birth weight equation, Example 4.9	例 4.9,出生体重方程中的父母受教育水平,146
parsimonious model	节省模型,194
partial derivative	偏导数,659
partial effect	偏效应,73,647
partial listing of data on U.S. inflation and unemployment rates, Table 10.1	表10.1,部分列举美国通货膨胀和失业率方面的数据,312
participation in 401(k) pension plans, Example 3.3	例 3.3,401k 养老金计划的参与,78
participation rates in 401(k) plans, Example 4.6	例 4.6,401k 计划的参与率,131
partitioned matrix multiplication	分块矩阵的乘法,748
percent correctly predicted	正确预测百分数,536
percentage change	百分比变化,649
percentage point change	百分点变化,650
percentages, proportions and	比例和百分比,648
perfect collinearity	完全共线性,84
permanent income hypothesis, testing, Example 16.7	例 16.7,对持久收入假说的检验,518
personal computer ownership, determinants of, Example 8.9	例 8.9,拥有个人计算机的决定因素,273
personal exemption on fertility rates, effects of, Example 10.4	例 10.4,个人税收豁免对生育率的影响,325
Philips curve, expectations augmented, Example 11.5	例 11.5,附加预期的菲利普斯曲线,357
plug-in solution to the omitted variables problem	遗漏变量问题的插入解,285
point forecast	点预测,596
Poisson distribution	泊松分布,547
Poisson regression for number of arrests, Example	例 17.3,拘捕次数的泊松回归,549

17.3	
Poisson regression model	泊松回归模型,546
police force, murder rates and size of, Example 16.1	例 16.1,谋杀率与警力规模,504
policy analysis	政策分析, 217
and program evaluation	~与计划评估, 238
with pooled cross sections	对混合横截面的~,414
with two-period panel data	对两期综列数据的~,426
pooled cross section	混合横截面, 10
and panel data, applying 2SLS to	将 2SLS 应用于~和综列数据, 488
policy analysis with	用~的政策分析,414
two years of housing prices, Table 1.4	表 1.4,两年的房屋价格的~, 11
pooled OLS using first differences, Assumption FD.1 - FD.7	利用一阶差分的混合 OLS,假定 FD.1 - FD.7, 439 440
pooling independent cross sections across time	不同时间独立横截面数据的混合, 409
population	总体, 699
model	~模型,32
R-squared	~*R*-平方,192
regression function(PRF)	~回归函数, 26
testing hypotheses about the mean in a normal	检验关于正态总体均值的假设, 725
positive definite	正定, 750
positive semi-definite	半正定, 750
power of a test	一项检验的功效, 725
practical significance	实际显著性,131,735
versus statistical	~对比统计显著性,735
Prais-Winsten (PW) estimation	普莱斯-温斯顿(PW)估计, 389
predetermined variable	前定变量, 517
predicted variable	预测变量, 23
prediction	预测, 197
and residual analysis	预测和残差分析,197
error	预测误差,200
interval	预测区间,200
predictor variable	预测元变量,23
prices, housing investment and, Example 10.7	例 10.7,住房投资和价格,335
prison population on violent crime rates, effect of, Example 16.8	例16.8,关押人数对暴力犯罪率的影响,521
probabilities for a normal random variable, Example B.6	例 B.6,正态随机变量的概率, 692
probability	概率
Figure B.2	图 B.2,668

limit	~极限,709
of success (response probability)	成功~(响应概率),233
probability density function (pdf)	概率密度函数, 666
Figure B.1	图 B.1,667
probability distribution	概率分布
features of	~的特性,672
Figure B.3	图 B.3,677
random variables and their	随机变量及其~,664
probit estimates, interpreting logit and	解释 logit 和 probit 估计值,535
probit models	probit 模型,531
for binary response, logit and	二值响应的 logit 和~,530
maximum likelihood estimation of logit and	logit 和~的最大似然估计, 533
specifying logit and	设定 logit 和~, 530
productivity, wages and, Example 11.7	例 11.7,工资与生产力, 365
program evaluation	项目评价, 217
proof of theorem	定理的证明
3.1 and 3.2	定理 3.1 和定理 3.2 的证明,110
3.4	定理 3.4 的证明, 111
property	性质
PLIM.1	PLIM.1, 710
PLIM.2	PLIM.2,711
some special functions and	一些特殊的函数和~, 650
sum.1 and sum.2	sum.1 与 sum.2, 643
sum.3	sum.3, 644
proportionate change	比例变化, 649
proportions and percentages	比例与百分比, 648
proxy variable	代理变量, 284
using lagged dependent variables as	用滞后因变量作为~, 289
pseudo R-squared	拟 R-平方,536
Puerto Rican employment	波多黎各就业
and the minimum wage, Example 10.3	例 10.3.~与最低工资, 324
Example 10.9	例 10.9, 338
Puerto Rican minimum wage, Example 12.7	例 12.7,波多黎各最低工资, 398
Q	
quadratic form	二次型,750
and positive definite matrices	~和正定矩阵,750
differentiation of linear and	线性和~的微分, 751
quadratic functions	二次函数, 185,651
quadratic relationship, Figure 6.1	图 6.1,二次关系,187

quadratic wage function, Example A.4	例 A.4,二次工资函数, 652
qualitative information	定性信息, 211
quantity, graph of, Figure A.2	图 A.2,数量图,648
quasi-differenced data	拟差分数据, 388
quasi-experiment	拟实验, 417
quasi-likelihood ratio statistic	拟似然比统计量,549
quasi-maximum likelihood estimation (QMLE)	拟最大似然估计(QMLE),548
question, posing	提出问题, 616
R	
R&D intensity	R&D 强度
against firm sales, scatterplot of, Figure 9.1	图 9.1,~相对于企业销售额的散点图, 302
and firm size, Example 9.8	例 9.8,~与公司规模,301
Example 9.9	例 9.9,302
R-bar squared	$\overline{R}^2$,193
R-squared	R-平方, 39
after IV estimation, computing	IV 法估计后计算~,471
form of the F statistic	F 统计量的~形式, 146
when the dependent variable is trending, computing	当因变量有趋势时,计算~,338
race discrimination in hiring	雇用中的种族歧视
Example C.3	例 C.3,723
Example C.5	例 C.5,730
Example C.7	例 C.7,734
random effects	随机效应
estimator	~估计量, 451
model	~模型, 449
or fixed effects	~或固定效应, 452
random sample	随机样本, 700
random sampling	随机抽样,67,699,700
Assumption MLR.2	假定 MLR.2, 82
Assumption SLR.2	假定 SLR.2, 47
Assumption 2SLS.2	假定 2SLS.2, 498
random variables	随机变量, 664
and their probability distribution	~及其概率分布,664
Figure B.4	图 B.4, 678
pairwise uncorrelated	两两不相关~, 683
standardizing	标准化~,679
variance of sums of	~和的方差, 682
random vector	随机向量, 751

random walk	随机游走，359
Figure 11.1	图 11.1,361
with drift	带漂移的～,361
Figure 11.3	图 11.3,363
rank	秩，749
of a matrix	一个矩阵的～，749
properties of	～的性质，750
rank condition	秩条件，480，510
Assumption 2SLS.4	假定 2SLS.4，499
for identification of a structural equation	识别一个结构方程的～，510
ratio, F	F 比率，142
rational distributed lag (18.16), lag distribution for, Figure 18.1	图18.1,有理分布滞后(18.16)的滞后分布，577
rational distributed lag (RDL) model	有理分布滞后(RDL)模型，576
reduced form	约简型,506
equation	～方程，474
error	～误差,506
parameters	～参数,506
regressand	回归子，23
regression analysis	回归分析
controlling for too many factors in	～中控制过多因素,196
using highly persistent time series in	～中利用高度持续性的时间序列，358
using trending variables in	～中利用趋势变量，334
regression functions across groups, testing for differences in	检验各组回归方程的(组间)差异，230
regression models	回归模型
examples of time series	时间序列～的例子，312
including irrelevant variables in	～中包含无关变量，87
used for forecasting, types of	用于预测的～类型，595
regression results, reporting	报告回归结果，150
regression specification error test(RESET)	回归设定误差检验(RESET),281
regression through the origin	过原点回归，58,81
regressions with a time trend, a detrending interpretation of	对含有时间趋势回归的除趋势解释,337
regressor	回归元,23
rejection region	拒绝域，726
against one-sided alternative, Figure C.5	图 C.5,相对于单侧对立假设的～,727
against two-sided alternative, Figure C.6	图 C.6,相对于双侧对立假设的～,729
rejection rule	拒绝法则,119
relationship between the probability of being in the	图7.3,所估计的、参与劳动市场和受教育年数

labor force and years of education, estimated, Figure 7.3	之间的关系,235
relative efficiency	相对有效性,707
relevant variable, excluding	排除有关变量, 87
RESET (regression specification error test)	RESET(回归设定误差检验), 281
as a general test for functional form misspecification	~作为对函数形式误设的一般检验, 281
residential price inflation, housing investment and, Example 18.1	例 18.1,住房投资与住房价格的上涨,577
residual analysis	残差分析, 201
prediction and	预测与~, 197
residual sum of squares (SSR)	残差平方和,38,78
residuals	残差,30
fitted values and	拟合值与~,36
Figure 2.4	图 2.4, 31
Table 2.2	表 2.2,36–37
minimizing the sum of squared	最小化~平方和, 65
OLS fitted values and	OLS 拟合值和~, 75
response probability (probability of success)	响应概率(成功概率),233
response variable	响应变量, 23
restricted model	受约束模型, 141
return on equity, CEO salary and	CEO 薪水与股票收益
Example 2.3	例 2.3,32
Example 2.6	例 2.6,36
Example 2.8	例 2.8,40
return to education	受教育的回报
and the gender wage gap, changes in, Examplel 3.2	例 13.2,~的变化与性别工资差异,412
changed over time, Example 14.2	例 14.2,~随时间变化,445
for married women, estimating, Example 15.1	例 15.1,估计已婚妇女~,467
for men, estimating, Example 15.2	例 15.2,估计男性~,468
measuring, Example 1.4	例 1.4,度量~,15
return to education for working women	工作妇女的教育回报
Example 15.5	例 15.5,478
Example 15.7	例 15.7,483
Example 15.8	例 15.8,485
revenue, finding expected, Example B.5	例 B.5,求期望收益,675
root mean squared error	根均方误,600
row vector,	行向量,745

S

salaries, effects of race on baseball players, Example 7.11	例 7.11,种族对棒球运动员薪水的影响,228
salary-benefits tradeoff, testing, Table 4.3	表 4.3,检验薪水与津贴之间的替换关系,152
salary-pension tradeoff for teachers, Example 4.10	例4.10,教师薪水与养老金之间的替换关系,151
sales tax increase (Michigan), Example A.3	例 A.3,(密歇根州)销售税的提高,650
sample average	样本平均, 701
sample correlation coefficient	样本相关系数, 714
sample covariance	样本协方差, 713
sample figure caption, Figure 17.2	图 17.2,样本图标,557
sample regression function (SRF)	样本回归函数, 31,72
sample selection correction	样本选择纠正,557,561
sample standard deviation	样本标准差,711
sample variance	样本方差, 704
sample variation in the independent variable, AssumptionSLR.4	假定 SLR.4,自变量的样本变异,49
sampling	抽样, 700
sampling distribution	抽样分布, 702
Figure 5.1	图 5.1,163
of a consistent estimator, Figure C.3	图 C.3,一个一致估计量的~,709
of the OLS estimators, Theorem 2.2	定理 2.2,OLS 估计量的~,55
of the OLS slope estimators, Theorem 3.2	定理 3.2,2OLS 斜率估计量的~,93
saving function with measurement error	含测量误差的储蓄方程,292
saving, housing expenditures and, Example 16.2	例 16.2,住房支出与储蓄,505
scalar multiplication	数乘, 746
scalar variance-covariance matrix	数量方差-协方差矩阵,759
scatterplot	散点图
R&D intensity against firm sales, Figure 9.1	图 9.1,R&D 强度相对于企业销售量的~,302
savings and income, and the population regression, Figure 2.2	图 2.2,储蓄与收入~和总体回归,28
wage against education, Figure 2.3	图 2.3,工资相对于受教育程度的~,30
school lunch program, student math performance and, Example 2.12	例 2.12,学生数学成绩与学校午餐项目,51
school size, student performance and, Example 4.2	例 4.2,学生的表现与学校规模,121
score statistic	得分统计量, 171
scrap rates	废弃率
effect of job training grants on firm, Example	例 4.7,在职(工作)培训津贴对企业~的影

4.7	响,132
effect of job training on firm, Example 14.1	例14.1,在职(工作)培训对企业~的影响,443
Example 14.3	例14.3,448
for 20 Michigan manufacturing firms, Table C.3	表C.3,密歇根20个制造企业的~,720-721
measurement error in, Example 9.6	例9.6,~中的测量误差,293
search services, on-line	在线搜索服务,619
seasonal dummy variables	季度虚拟变量,340
seasonality	季节性, 340
trends and	趋势与~,331
seasonally adjusted	季节调整,340
selected sample	选择样本,558
self-selection	自选择, 239
semi-elasticity	半弹性,46,183,188,190,656
sensitivity analysis	敏感分析,624
serial correlation	序列相关, 321
correcting for higher order	对高阶~的修正, 393
differencing and	差分与~,394
dynamically complete models and the absence of	动态完整模型与不存在~,366
in regression models, heteroskedasticity and	回归模型中的~和异方差性, 402
in the presence of lagged dependent variables	出现滞后因变量时的~,378
testing for	对~的检验,380
higher order	高阶,385
with general regressions, testing for	检验一般回归的~,384
with strictly exogenous regressors, correcting for	修正具有严格外生回归元的~, 387
serial correlation-robust inference after OLS	OLS后序列相关-稳健的推断,395
serial correlation-robust standard error	序列相关-稳健的标准误,396
for $\hat{\beta}_j$	$\hat{\beta}_j$~,397
serially correlated errors, properties of OLS with	具有序列相关的OLS的性质,376
serially uncorrelated	序列无关,366
shifting supply equations traces out the demand equation, Figure 16.1	图16.1,移动供给方程描出需求方程,509
short-run elasticity	短期弹性,325
significance, economic or practical versus statistical	经济或实际与统计显著性,131,735
significance level	显著性水平, 119,725
simple linear regression model	简单线性回归模型,22
simple regression	简单回归

incorporating nonlinearities in	～中包含非线性关系,42
terminology for, Table 2.1	表2.1,～术语表,23
simple regression model	简单回归模型,22
omitted variables in	～中的遗漏变量,462
under homoskedasticity, Figure 2.8	图2.8,同方差性下的～,53
simple rule of thumb for a 95% confidence interval	95%的置信区间的简单经验法则,772
simple wage equation, Example 2.2	例2.2,简单工资方程,24
simulated confidence intervals, from a normal distribution, Table C.2	表C.2,由正态分布模拟的置信区间,717－718
simulation of estimators for a normal distribution, Table C.1	表C.1,模拟正态分布的估计量,706－707
simultaneity	联立性,501
bias	～偏误,507
in OLS	OLS中的～506
simultaneous equations models(SEMs)	联立方程模型.501
nature of	～的性质,501
with panel data	含有综列数据的～,520
with time series	含有时间序列的～,516
single dummy independent variable	单虚拟自变量,213
single endogenous explanatory variable	单内生解释变量,476
single explanatory variable, testing for a(n) endogeneity of	检验单解释变量的内生性,484
single linear combination of the parameters, testing hypotheses about	检验参数单一线性组合的假设,136
sleeping versus working, Example 13.5	例13.5,睡眠与工作,424
slope	斜率,646
allowing for different	考虑不同～,226
parameter	～参数,23,69
soybean yield and fertilizer, Example 2.1	例2.1,大豆产量与肥料.24
special form of the White test in the log housing price equation, Example 8.5	例8.5,对数住房价格方程中特殊形式的怀特检验,260
specification issues in Tobit modelsTobit	模型中的设定问题,545
spreadsheet	空白表格程序,621
spurious regression	谬误回归,335,584
problem	～问题,585
square matrix	方阵,744
stable AR(1) process	稳定的AR(1)过程,350
standard deviation	标准差,676,679

of $\hat{\beta}_j$	$\hat{\beta}_j$ 的～，100
property SD.1 and SD.2	～的性质 SD.1 和 SD.2,679
standard error	标准误，720
in a birth weight equation, Example 5.2	例 5.2,出生体重方程中的～，170
of $\hat{\beta}_1$	$\hat{\beta}_1$ 的～,58
of $\hat{\beta}_j$	$\hat{\beta}_j$ 的～,100
of the regression (SER)	回归～,58,100
on the OLS estimators, estimating σ^2	OLS 估计量的～:估计 σ^2,99
standard normal cumulative distribution function, Figure B.8	图 B.8,标准正态累积分布函数,690
property normal.1	～的性质 normal.1，691
property normal.2, normal.3, and normal.4	性质 normal.2,normal.3 和 normal.4，692
standardized coefficients	标准化系数，182
standardized random variable	标准化随机变量，680
state infant mortality rates, Example 9.10	例 9.10,州婴儿死亡率,303
static model	静态模型,313
Example 11.1	例 11.1,353
static Phillips curve	静态菲利普斯曲线
Example 10.1	例 10.1，323
Example 12.5	例 12.5，392
stationary and nonstationary time series	平稳和非平稳时间序列，348
stationary and weakly dependent time series	平稳和弱依赖时间序列，347
stationary process	平稳过程，348
stationary stochastic process	平稳随机过程，348
statistical inference	统计推断，761
with the IV estimator	有 IV 估计量的～，465
statistical significance	统计显著性
economic or practical versus	经济或实际与～，131
practical versus	实际与～,735
statistically insignificant	统计不显著,125
statistically significant	统计显著，125
stingily dependent	强相依，359
stochastic process	随机过程，312
stock returns, ARCH in, Example 12.9	例 12.9,股票回报中的 ARCH,402
strict exogeneity	严格外生性，422
strict exogenous	严格外生，318
structural equation	结构方程，472，502
identifying and estimating	～的识别和估计，507
structural errors	结构误差，503
structural parameters	结构参数，506

student performance and school size, Example 4.2	例 4.2,学生表现与学校规模,121
sum of squared residuals (SSR)	残差平方和, 30,78
minimizing	最小化～, 65
summary of bias, Table 3.2	表 3.2,偏误的总结,90
summary of functional forms involving logarithms, Table 2.3	表 2.3,涉及对数的函数形式总结, 45
summary statistics, Table 2	表 2,摘要统计量, 633
summation operator	求和运算子, 643
symmetric matrix	对称矩阵, 748
symmetric probability distribution, Figure B.3	图 B.3,对称概率分布, 677
systems with more than two equations	两个以上方程的方程组, 514
systems with three or more equations, identification in	两个以上方程的方程组中的识别,514
T	
t distribution	t 分布,693,753
Figure B.10	图 B.10, 695
Figure C.4	图 C.4,719
t ratio	t 比率, 117
t statistic	t 统计量, 117,727
t tests	t 检验
computing p-values for	计算 p 值或～,128
for AR(1) serial correlation with strictly exogenous regressors	对含有严格外生回归元的 AR(1)序列相关～,380
testing hypotheses about a single population parameter	检验单一个总体参数假设的～,116
T-bill rate, Figure 11.2	图 11.2,国库券利率, 362
unit root test for three-month, Example 18.2	例 18.2,对 3 个月期～的单位根检验, 580
terminology	术语
for multiple regression, Table 3.1	表 3.1,多元回归的～, 70
for simple regression, Table 2.1	表 2.1,简单回归的～, 23
note on	对～的注解, 35
test consistency	检验一致性, 736
test statistic	检验统计量, 726
text (ASCII) file	文本文件, 620
text editor	文本编辑, 621
Theorem 15A.1 – 15A.3	定理 15A.1 – 15A.3, 499
Theorem E.6	定理 E.6, 762
three different estimators of a wage equation, Table 14.2	表 14.2,一个工资方程的三个不同估计量,452

time index	时间指标，316
time series	时间序列
characterizing trending	刻画趋势的～，331
data	～数据，8，311
equations, applying 2SLS to	在～方程中应用 2SLS，486
is I(1), deciding whether	决定～是否 I(1)，364
models, homoskedasticity assumption for	～模型的同方差性假定，369
nonstationary	非平稳～，348
process	～过程，312
regression models, examples of	～回归模型的例子，312
regressions, heteroskedasticity in	～回归中的异方差性，398
simultaneous equations models with	含有～的联立方程模型，516
stationary	平稳～，347，348
weakly dependent	弱相依～，347，349
time trend	时间趋势，331
time-demeaned	除去时间均值后的，442
time-varying error	随时间变化的误差，420
Tobit estimates, interpreting	解释 Tobit 估计值，541
Tobit estimation of annual hours worked, OLS and, Table 17.2	表17.2，年工作小时数的 OLS 和 Tobit 估计，544
Tobit model	Tobit 模型，540
specification issues in	～中的设定问题，545
top coding	顶端编码，552
total sample variation	总样本变异，94
total sum of squares (SST)	总平方和，38，78
trace	迹，748
of a matrix	一个矩阵的～，748
properties of	～的性质，749
traffic fatalities, effect of drunk driving laws on, Example 13.7	例 13.7，酒后驾驶法律对交通致死的影响，428
training grants and worker productivity, Example C.8	例 C.8，培训津贴与工人生产力，735
training grants on hours of training, effects of, Example 7.3	例 7.3，培训津贴对培训小时数的影响，217
transpose	转置，747
treatment group	处理组，217
trend-stationary process	趋势平稳过程，351
trending variables in regression analysis, using	利用回归分析中的趋势变量，334
trends and seasonality	趋势和季节性，331
true model	真实模型，82

truncated regression model	断尾回归模型，551，555
censored and	截取和～，551
two stage least squares (2SLS or TSLS)	两阶段最小二乘(2SLS 或 TSLS)，461,476
assumption for	～的假定，498
estimation by	由～估计，512
estimation, testing multiple hypotheses after	检验～估计后的多重假设，480
multicolinearity and	多重共线性与～，479
testing for AR(1) serial correlation after	检验～后 AR(1)序列相关，487
to pooled cross sections and panel data, applying	用于混合横截面和综列数据,488
to time series equations, applying	在时间序列方程中应用～，486
with AR(1) errors	含 AR(1)误差的～，488
with heteroskedasticity	含异方差的～，486
two test scores as indicators of ability, using, Example 15.6	例15.6,用两个测验得分作为能力的标示变量，482
two-equation system, identification in	两方程系统中的识别，508
two-period panel data	两期综列数据
analysis	～分析，419
policy analysis with	用～做政策分析，426
two-sided alternative	双侧对立假设，123，726
rejection region against, Figure C.6	图 C.6,相对于～的拒绝域，729
two-tailed test	双尾检验，124，728
two-variable linear regression model	双变量线性回归模型，22
two-year panel data set on city crime statistics, Table 1.5	表 1.5,城市犯罪统计的两年综列数据集，12
Type I error	第 I 类错误，725
Type II error	第 II 类错误，725

U

unbalanced panel	不平衡综列数据，448
unbiased estimation of σ^2	σ^2 的无偏估计
Theorem 10.3	定理 10.3，322
Theorem 2.3	定理 2.3，57
Theorem 3.3	定理 3.3，100
unbiased estimator	无偏估计量,702
Figure C.1	图 C.1,703
unbiasedness	无偏性，702，703
and consistency	～和一致性,376
of σ^2, Theorem E.4	定理 E.4,σ^2 的～，761
unbiasedness of OLS	OLS 的无偏性，46，316
Theorem 10.1	定理 10.1，319

Theorem 2.1	定理 2.1, 50
Theorem 3.1	定理 3.1, 86
Theorem E.1	定理 E.1, 758
unconditional forecast	无条件预测, 595
underspecifying the model	模型设定不足, 87
unemployment	失业
claims, effect of enterprise zones on, Example 13.8	例 13.8,企业特区对～补贴申请的影响,431
effect of the minimum wage on, Example 1.6	例 1.6,最低工资对 - 的影响, 17
forecasts, out-of-sample comparisons of, Example 18.9	例 18.9,～样本外预测比较, 600
unemployment rates	失业率
city, Example C.1	例 C.1,城市～, 701 - 702
forecasting the U.S., Example 18.8	例 18.8,预测美国～. 596
partial listing of data on U.S. inflation and, Table 10.1	表 10.1 部分列举美国通货膨胀和～数据, 312
two-year-ahead forecast for, Example 18.10	例 18.10,对失业率提前两年的预测, 603
unidentified equation	不可识别方程, 515
unit root	单位根, 578
in the log of U.S. real gross domestic product, Example 18.4	例 18.4,美国真实 GDP 对数中的～,583
process	～过程, 360
test for annual U.S. inflation, Example 18.3	例 18.3,对美国年度通货膨胀的～检验, 582
test for three-month, T-bill rates, Example 18.2	例 18.2,对 3 个月期国库券利率的～检验, 580
unit root t test	单位根 t 检验
linear time trend, asymptotic critical values for, Table 18.3	表18.3,线性时间趋势下～的渐近临界值, 583
no time trend, asymptotic critical values for, Table 18.2	表 18.2,无时间趋势下～的渐近临界值,580
unobserved effect	不可观测效应, 420
unobserved effect model	不可观测效应模型, 420
unobserved heterogeneity	不可观测的异质性, 420
unrestricted model	无约束模型, 141
upward bias	向上偏误, 91

V

value, computing an expected, Example B.3	例 B.3,计算一个期望值, 673

Var($\hat{\beta}_1$) as a function of R_1^2, Figure 3.1	图 3.1,作为 R_1^2 的一个函数的 Var($\hat{\beta}_1$), 96
Var(*wage*|*educ*) increasing with *educ*, Figure 2.9	图 2.9,Var(*wage*|*educ*)是 *educ* 的增函数, 54
variability, measures of	变异性的度量, 676
variable	变量
binary	二值~, 211
control	控制~, 23
dependent	因~,23
descriptions, Table 1	表 1 描述~,632-633
dummy	虚拟~, 211, 220
endogenous explanatory	内生解释~, 84, 278
exogenous explanatory	外生解释~, 84
explained	被解释~,23
explanatory	解释~, 23
inclusion of an irrelevant	包含一个无关~, 87
independent	自变量, 23
model with *k* independent	含有 *k* 个自~的模型, 69
model with two independent	含有 2 个自~的模型, 66
ordinal	序数~, 222
predicted	预测~23,
predictor	预测元~,23
response	响应~, 23
single dummy independent	单一个虚拟自~, 213
zero-one	0-1~, 211
variance	方差, 676, 677
error	误差~,53,94
estimating the error	估计误差~, 56
in misspecified models	误设模型中的~, 97
of sums of random variables	随机变量和的~, 682
of the OLS estimators	OLS 估计量的~, 52, 92, 320
and the Gauss-Markov theorem	~和高斯-马尔科夫定理, 320
of the prediction error	预测误差的~, 200
OLS sampling, Theorem 10.2	定理 10.2,OLS 抽样方差 ,321
properties of	~的性质, 752
property VAR.1 and VAR.2	~的性质 VAR.1 和 VAR.2, 678
property VAR.3 and VAR.4	~的性质 VAR.3 和 VAR.4, 683
variance-covariance matrix	方差-协方差矩阵, 752
of the OLS estimator, Theorem E.2	定理 E.2,OLS 估计量的~, 759
vector	向量, 745
autoregressive (VAR) model	~自回归(VAR)模型, 598

column	列～，745
row	行～，745
voting outcomes and campaign expenditures	投票结果与竞选支出
Example 2.5	例 2.5，34
Example 2.9	例 2.9，41
W	
wage	工资
and education	～与受教育水平
Example 2.4	例 2.4，34
Example 2.7	例 2.7，38
and other individual characteristics, cross-sectional data set on, Table 1.1	～表 1.1，～与其他个人特征的横截面数据集，7
and productivity, Example 11.7	例 11.7，～与生产力，365
effects of computer usage on, Example 7.9	例 7.9，计算机使用对～的影响，225
effects of physical attractiveness on, Example 7.7	例 7.7，相貌吸引力对～的影响，222
Figure 2.6	图 2.6，43
function with interaction, Example A.8	例 A.8，含有交互项的～函数，660
wage equation	工资方程
heteroskedasticity in, Example 2.13	例 2.13，～中的异方差性，54
three different estimators of, Table 14.2	表 14.2，～的三个不同估计量，452
using panel data, Example 14.4	例 14.4，利用综列数据的～，451
wage offer equation for married women	已婚妇女的工资报价方程
Example 17.5	例 17.5，562
Table 17.5	表 17.5，563
weakly dependent	弱相依，349
time series	～时间序列，349
weighted least squares	加权最小二乘，399
estimating the linear probability model by	用～估计线性概率模型，272
estimation	～估计，261
(WLS) estimators	～估计量，263
White test for heteroskedasticity	异方差性的怀特检验，259
special case of	～的特殊情形，260
White test in the log housing price equation, special form of, Example 8.5	例8.5，对数住房价格方程中怀特检验的特殊情形，260
within estimator	组内估计量，442
within transformation	组内变换，442
women's fertility, determinants of, Table 13.1	表 13.1，妇女生育能力的决定因素，411
women's fertility over time, Example 13.1	例 13.1，不同时期的妇女生育能力，409

worker compensation laws on duration, effect of, Example 13.4	例13.4,劳工救济法对持续期(久期)的影响,418
worker productivity	工人生产力(率)
effect of job training grants on, Example C.2	例C.2,在职(工作)培训津贴对～的影响,720
Example C.6	例 C.6, 732
job training and, Example 1.2	例 1.2,在职(工作)培训与～, 4
Example 15.10	例 15.10, 489
training grants and, Example C.8	例 C.8,培训津贴与～,735
working, sleeping versus, Example 13.5	例 13.5,睡眠与工作, 424

Y

year dummy variables	年度虚拟变量, 409

Z

zero conditional mean assumption	零条件均值假定, 25
Assumption MLR.3	假定 MLR.3, 83
Assumption SLR.3	假定 SLR.3, 47
Assumption TS.2	假定 TS.2, 317
Assumption TS.2′	假定 TS.2′,352
zero matrix	零矩阵, 745
zero mean and zero correlation, Assumption MLR.3′	假定 MLR.3′,零均值和零相关,164
zero-one variable	0－1 变量, 211

译后记

计量经济学在20世纪后半叶的发展，从根本上影响着经济学研究的方法与进展。1969年首届诺贝尔经济学奖授予两位宏观计量经济学家这一事实，已经标志着计量经济学在经济学的理论和经验研究中具有举足轻重的地位；而J. Heckman和D. McFadden两位微观计量经济学家又在2000年获得诺贝尔奖，再次表明计量经济学在整个经济学体系中不可或缺的地位。在2000年初版的这本书，不仅见证了计量经济学在这半个世纪的发展，涵盖了该学科几乎所有的经典理论，而且还用简洁准确的语言阐述了计量经济学研究的最新进展。与其他计量经济学教科书相比，本书最重要的特点是：更便于学生对计量经济学的理解和运用；更有利于指导经济研究者在经验研究中使用计量经济学的方法。所以，本书是对传统计量经济学教学和研究的一种突破（或许正是这个原因，使古扎拉蒂的《计量经济学》紧随其后大幅调整体系而出版了第4版）。

年逾八旬的林少宫教授刚完成古扎拉蒂的《计量经济学》和约翰斯顿的《计量经济学方法》两本书的翻译，原本打算休息一段时间，但在看到这本书后，立即觉得它是一本更适合于学习和研究使用的教材，并坚持要找人协助他尽早把这本优秀的教科书奉献给国内的读者。加之美国佛罗里达大学计量经济学副教授艾春荣博士在访问华中科技大学时，也对此书极为推崇，并声称是从本书作者那里学习计量经济学的，这就更加坚定了林老的信念。于是林老便在2002年1月通过考试的方式甄选翻译人员，并在2002年7月初完成了全部译稿。林老为此付出了艰辛的劳动和大量的时间，不仅在初译时逐字逐句地把关，而且在全部书稿完成后，又用近一个月的时间进行最后审校，以使读者尽可能正确而又顺利地接触原著的思想。

在此中文版完成之际，原著的第2版也即将面世（2003），但从第2版的内容、特点和更新的

篇幅来看，都不值得读者等那么长的时间去读一本好书，尤其是与本书相对其他计量经济学的进步而言，这种重版带给人的新的感觉更是微不足道的。所以，我们决定立即将这已经译好的第1版本献给读者，对第2版感兴趣的读者，可从本书提供的网址查阅其英文版目录及特色介绍。

本书是在林老的组织和审定下完成的，林老还亲自翻译了前言、第13章、第14章及附录A，B，C。林相森翻译了全书的时间序列部分，即第10～12章和18章；研究生肖遥翻译了第2章和19章；盛抒扬翻译了第15章；其余部分由费剑平翻译。

感谢华中科技大学经济学院的张培刚教授、徐长生教授、张卫东教授、方齐云教授、张建华教授对我在工作中因不得已延误了一些时间的宽容。感谢本书策划梁晶工作室和责任编辑所付出的努力，他们使本书的质量有显著提高，并使本书尽快尽好地呈现在读者面前。也感谢我妻子孙春霞承担了我在儿子出生到半岁期间理应承担的许多劳动，使我有足够完整的时间去做这件有意义的工作。

尽管林老和责任编辑付出了相当的努力，但文中仍难免有错误和不当之处，对此译者责无旁贷。

费剑平
于武汉东湖
2002年7月